1991

The Role Of Natural Constraints In Computational Theories Of Vision

Peter Alan Morton

THE ROLE OF NATURAL CONSTRAINTS IN COMPUTATIONAL THEORIES OF VISION

by

Peter A. Morton

Department of Philosophy

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
December 1990

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

ISBN 0-315-64285-8

ABSTRACT

The thesis examines the philosophical implications of the computational theory of early vision developed by Marr. According to Marr, early visual processes consist of sequences of "modular" computational mechanisms. These processes rely on functional relations between rates of change in stimulus magnitudes which result from certain contingent, global properties -- natural constraints -- of the physical world.

Marr argues that explanations of early vision must have three distinct levels of description: computational, algorithmic and physical. In Chapter 1 I defend the explanatory significance of this distinction in levels. In fulfilling its role in describing the dependence of visual processes on natural constraints, the computational level forms an autonomous level of description in the sense that it is unaffected by the computational steps at the other levels.

In Chapter 2 I discuss the implications of natural constraints for the issues of individualism and methodological solipsism. I conclude that Marr's theory is nonindividualistic in the sense that visual content does not supervene on neural properties. However, this merely reflects the fact that different computational theories may be selected for the

same system. Importantly, Marr's theory does not violate methodological solipsism since interpretations within theories must supervene on neural properties.

In Chapter 3 I argue from the results of Chapters 1 and 2 that psychological explanation does not reduce to neurophysiology. This conclusion does not follow from the functionalist argument against physicalism, which is based on an incorrect account of computational theories. Rather the conclusion reflects the explanatory incompleteness of neurophysiological theories given the autonomous role of the computational level.

In Chapter 4 I look in detail at the arguments for a "language of thought" as they apply to early vision. I distinguish two versions of the language of thought hypothesis, one weaker than the other. I conclude that the stronger version, which claims that a cognitive system is "program-using", is false of early vision because of the role of natural constraints. The weaker claim that cognitive processes employ symbolic transformations is true of the computational-level theory of early vision, but there is insufficient evidence to establish the claim at the algorithmic level.

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to William Demopoulos. I am grateful to Bill for his encouragement and patience as the thesis developed. I have also benefitted a great deal from ideas in several of his essays, and from many suggestions in conversation.

I also wish to thank Keith Humphrey for many helpful conversations, and for his assistance in searching through the psychological literature. I should mention too that the ideas developed here began initially with an insightful remark of Keith's in conversation.

Philip Catton helped me with the technical discussions; I am grateful for his assistance.

And finally, I would like to extend a note of thanks to Robert Butts for his encouragement in my studies.

TABLE OF CONTENTS

Page

CERTIFICATE OF EXAMINATION ii
ABSTRACT iii
ACKNOWLEDGEMENTS v
TABLE OF CONTENTS vi
LIST OF APPENDICES viii

CHAPTER I -- NATURAL CONSTRAINTS
AND THE COMPUTATIONAL LEVEL 1

0. Introduction 1
1. The basic levels 2
2. Early vision:
a sketch of the explanatory framework 11
3. Decomposition at the computational level 15
4. Summary 25

CHAPTER II -- SUPERVENIENCE AND COMPUTATIONAL
EXPLANATION 27

0. Introduction 27
1. The "Argument from Success" 30
2. Assessment of the Argument from Success 35
3. Supervenience and
methodological solipsism 40
4. Summary 48

CHAPTER III -- AUTONOMY AND REDUCTION OF
PSYCHOLOGICAL THEORIES 50

0. Introduction 50
1. Reduction and physicalism 52
2. Kitcher's argument from
explanatory incompleteness 56
3. Reduction and autonomy 60
4. Optimality and idealization in
computational-level theory 65
5. Summary 69

CHAPTER IV -- MODULAR PROCESSES AND THE LANGUAGE OF THOUGHT 71

0. Introductien 71
1. The language of thought hypothesis 72
2. One argument from nonprojectible properties 81
3. A better argument from nonprojectible properties and its application to early vision 88
4. Functional architecture and symbolic transformations 95
5. A language of thought at the computational level 98
6. Why a language of thought? 102
7. Summary 105

APPENDIX I 107

APPENDIX II 114

APPENDIX III 127

APPENDIX IV 142

BIBLIOGRAPHY 156

VITA 160

LIST OF APPENDICES

Appendix	Page
APPENDIX I GIBSON'S THEORY OF DIRECT PERCEPTION	107
APPENDIX II STEREOPSIS AND THE MATCHING PROBLEM ..	114
APPENDIX III NEUROPHYSIOLOGY AND IMAGE INTERPRETATION	127
APPENDIX IV THE ROLE OF EMPIRICISM IN HELMHOLTZ'S THEORY OF VISION	142

CHAPTER ONE

NATURAL CONSTRAINTS AND THE COMPUTATIONAL LEVEL

0. Introduction.

The general purpose of this chapter is to argue that Marr's distinctions between computational, algorithmic and physical levels of description are of fundamental importance in understanding early visual processes. (See Marr 1982 and the papers cited there.) I believe that my articulation of the distinctions is faithful to Marr's understanding of the issues, but my primary objective is not exegetical. The central conclusion of my argument is that there is a distinguished computational level of decomposition of early visual processes that has a unique explanatory role. Explanation at the computational level consists of a sequence of functions that yield successive representations of the stimulus. The primitive functions of the sequence are postulated to explain the dependence of successful perception on contingent facts about the world. Thus in important respects the explanatory role of the computational level is the same as that of Gibson's theory of ecological optics, discussed in Section 3 of this chapter and in Appendix 1. In fulfilling this explanatory role, the computational level forms an "autonomous" level of decomposition in the sense that the sequence of functions in the computational-level description of early vision is not determined by the steps in the algorithmic- and physical-level decompositions of visual processes.

1. The basic levels.

In this section I give a preliminary discussion of the distinctions between the computational and algorithmic levels of description, and between the algorithmic and physical levels. The intention is to gain an introductory understanding of the levels as Marr presents them.

The levels of description were devised by Marr primarily as an articulation of computationalism in perception theory: the assertion that the neurological processes that underlie perceptual abilities are in some sense computing devices. But Marr intends his distinction between levels as more than simply a tool for modelling the visual system. The levels of description are viewed as having broad significance for understanding computing mechanisms generally. In this section we will see that in its widest application the distinction between computational and algorithmic levels is relativized to a context of enquiry. On such a reading, a description of a function which forms part of a computational-level theory in one context may in another context form part of an algorithmic-level decomposition of the same system. Our first task is to establish a preliminary outline of the distinctions.

Let us consider what it means to say that a physical mechanism is a computing device. A mechanism M is given a computational description by specifying an interpretation function I from a distinguished set of physical states to the union of the domain and range of a function F, the function "computed" by the system. Let us call the set C of states of M in the domain of I the set of computational states of M under I. A mechanism M is said to compute F if, in virtue of the natural laws that govern its behaviour, there is a physical-state transition map P:C-->C such that, for c ε C,

$$I(P(c)) = F(I(c)).$$ [1]

According to Marr, a complete explanation of the operation of a computing device must be based on a clear description of the purpose it is designed (or has evolved) to serve. The role of the computational level of theory is to relate the computational behaviour of M to this description of its intended purpose. Since the purpose of any computing device is the mechanical extraction of information,

[1]. As Stabler (1987) and Lycan (1981) point out, a difficulty in explicating the nature of computational theories is that under some interpretation every system computes every function. For example, it is possible to construct trivial computational theories where the appropriate computational relation among physical states is achieved by building the function to be computed into the interpretation function. This is a difficult problem whose solution will require some account of the explanatory roles of computational theories.

such a description will include the information that is required from the operation and the information that is available to it. A computational-level description of M has two parts: (1) a description of the function F computed by M; (2) a demonstration that the computation of F yields the information required from the information that is available.

In certain cases the states in C are simple physical magnitudes or discrete physical states of M.[2] More commonly, the states in C are complex states that correspond to sets of symbols, and the state-transition map P is a computation of a symbolic transformation function. The algorithmic level of theory describes the manner in which a function is computed by a particular symbolic transformation. An algorithmic-level description has two parts: (1) a description of what Marr calls a "formal scheme" that describes how the domain and range of F is "encoded" by sets of symbols; (2) an "algorithm" consisting of a sequence of functions that describes how the appropriate output is generated by symbolic transformations from each set of input symbols. Let us see what this means.

---

[2]. For example, Fodor and Block (1973) describe such a mechanism that computes the value of m/n, consisting of a power source, a variable resistor and an ammeter, all connected in series; if the voltage is m volts and the resistance is n ohms, then by Ohm's Law the current is m/n amperes.

We can think of a formal scheme for a function F as consisting of (1) an "encoding function" E that maps sets of symbols to the union of the domain and range of F, and (2) a "symbolic transformation function" T:S-->S, where S denotes the symbol sets in the domain of E, such that

for s ε S,

$$E(T(s)) = F(E(s)).$$

I will refer to a set of symbols in the domain of an encoding function as a symbolic representation of the objects or states of affairs to which it is mapped.

Marr's notion of an algorithm corresponds to the standard account in theory of computation. On this account, an algorithm for a function F is a sequence $<F_1, F_2, \ldots, F_n>$ of functions where the value of each function F_i, for $1 \leq i < n$, is the argument for F_{i+1}, and such that for any argument x in the domain of F,

$$F(x) = F_n(F_{n-1}(\ldots F_1(x)).$$

Typically an algorithm will be a description of a sequence of steps in a symbolic transformation that "realizes" F. By this I mean that for each F_i in the algorithmic sequence there is a formal scheme for F_i consisting of an encoding function E_i, and a symbolic transformation function T_i such that the sequence $<T_1, T_2, \ldots, T_n>$ produces the same symbolic transformation as the function T.

The physical level of theory describes how a symbolic transformation function is computed by a mechanism M. A physical-level description consists of (1) a specification of how the computational states of M correspond to symbol sets in the domain of an encoding function E, and (2) a description of how a symbolic transformation function T is computed by M. We satisfy (1) and (2) by showing that the interpretation function I maps the computational states in C to sets of symbols in the domain of E, and that the physical-state transition function P is such that,

$$I(P(c)) = T(I(c)).$$

Although specification of the formal scheme realized in a mechanism is a question at the algorithmic level, we can also see that for reasons of rigour the computational level description must describe the function as taking one set of symbolic representations to another. If it is to be adequate, the computational level must demonstrate that all of the information required is obtainable by mechanical operations on physical states. The difference between the appeal to formal schemes at the two levels lies in the explanatory purposes they fulfil. The computational scheme must be a canonical description of the computation performed, designed to demonstrate the completeness and accuracy of the extraction of information. No claim is made that the description of the formal scheme is realized in the physical structure of the mechanism. We can think of the computational-level description as specifying a "virtual mechanism" that characterizes a class of devices that perform the same information-processing operation. The algorithmic-level description identifies which scheme is a true description of the implementation of the operation in a particular physical device.

Marr's emphasis on the description of a "formal scheme" suggests that he views the algorithmic level in the manner of Pylyshyn, as requiring the characterization of a canonical language for the system.[3] Despite this appearance there is no deep correspondence between the two notions of an algorithmic level, since the explanatory roles of an algorithmic level are completely different in the two accounts. Although this issue will be discussed at length in Chapter 4, it is instructive to compare Marr and Pylyshyn in some detail here.

According to Pylyshyn, computational models of cognitive processes must do more than describe the input-output behaviour of these processes; they must also describe how the behaviour is realized in the primitive computational operations of the system. This much is in accord with the aims of Marr's algorithmic level. The difference lies in Pylyshyn's understanding of primitive computational mechanisms. Pylyshyn refers to the task of describing the primitive operations of

---

[3]. Pylyshyn (1984). See also Fodor (1975 and 1987).

the system as that of modelling its "functional architecture". Pylyshyn's concern in developing the idea of functional architecture is to make precise the claim that computational descriptions capture behavioural generalizations that cannot be formulated in neurological theories, since the generalizations involve regularities in the behaviour of the nervous system that have no finitary description except in terms of what the states of the system represent. On Pylyshyn's view, the only processes that exhibit these regularities are those that have an internal syntax in the sense that the set of states in their memory is recursively generable. (The premise for this is that the input-output behaviour of mechanisms without an internal syntax can be explained entirely in terms of the physical or neurological processes that instantiate the state transition map.) The functional architecture of the system consists of those computational operations that do not themselves require an internal syntax. According to Pylyshyn, characterizing a process of the functional architecture is a matter of describing its formal properties in such a way that a complete description of the functional architecture of a system coincides with a canonical description of the "language of thought".

None of these concerns is involved in Marr's justification of an algorithmic level. The idea of a "formal scheme" is not intended to yield a characterization of a "language of thought"; it is meant only to reflect the fact that different physical systems may instantiate the same symbolic encoding of a function. Nor is an algorithm, in Marr's sense, tied to a distinguished set of primitive operations in the manner suggested by Pylyshyn.

Let me end this section on a general note about the levels of description. In theory of computation the difference between computational and algorithmic levels of description is just that between a function and its algorithmic decomposition. Hence in the account of an algorithmic level that we find within theory of computation there is no sense in which, outside of a context of enquiry, a particular function in a complex operation is part of a computational-level or an algorithmic-level description of the system. The particular level to which a function belongs depends on the process whose explanation we are seeking. Within theory of computation the question whether a function, computed by a mechanism M, is part of a computational or algorithmic level description of M is thus relativized to a context of enquiry.

The same is true of Marr's general formulation of the levels of description. What is peculiar to Marr's notion of a computational level of theory is his emphasis on the purpose of a computation. But a function in an algorithmic-level decomposition may have its own computational-level theory in this sense. In developing an algorithm we may know what information is required at a particular stage of the algorithm without knowing how that information may be extracted from some earlier stage. Specifying an information-processing goal may thus form part of explaining how an algorithm computes a complex function.

Neither is the claim that a computation is a realization of a formal scheme inconsistent with the assertion that the distinction between computational and algorithmic levels is context-relative. A description of a set of symbolic transformations that best captures generalizations at one level of decomposition is not thereby the best description at other levels of decomposition. This is revealed, for example, in the distinction between programming languages and machine languages. So while a set of symbolic transformations may constitute an algorithmic decomposition of an information-processing function at a given level of organization in the system, the primitive operations in that set may have a decomposition into another algorithmic sequence.

Hence on Marr's broad account of the levels, the distinction between computational- and algorithmic-level descriptions of a complex system is context-relative. I have raised this point because the same is not true of Marr's application of the two levels in his empirical theories of early vision. In Marr's empirical theories there is a distinguished sequence of functions that is said to comprise a computational-level description of the visual system. Hence the use Marr makes of the computational and algorithmic levels in theories of vision is not based solely on the general formulation of the distinction. Let us begin with an outline of the sequence of functions which, according to Marr's theory of vision, constitutes a computational-level description of the visual system.

2. Early vision: a sketch of the explanatory framework.

The basis of Marr's explanations of visual abilities is the fact that a great deal of information about the spatial layout of the distal scene can be obtained by computational processes that are, in his words, "modular"; i.e., they are relatively "independent" in that each has access to a limited and specific range of information.[4] According to the theories of Marr, initial perception of the distal scene is computed by a collection of such modular processes, where the computational behaviour of each is independent of the output of the others, and is not influenced by the output of "higher" cognitive

---

[4]. It would be better to use "autonomous" here rather than "independent", which is Marr's term. I have followed Marr's use because I have already reserved "autonomous" to describe the computational level.

processes.[5] According to Marr, the modular processes of early vision compute a sequence of functions that yield as output successive representations of the stimulus culminating in a representation of the three-dimensional shape and layout of objects in the field of view. A description of this sequence of functions, and of the representations generated by each function in the sequence, comprises what Marr refers to as the computational level of the theory of early visual processes.

Representations generated by the initial functions in the sequence make explicit the shapes and patterns in the spatial arrangement of light intensity levels in the retinal image that carry information about the physical world. The first representation of the image, referred to as the grey level image, represents only local-receptor intensity levels at each point on the retina. The second representation of the image, which Marr calls the raw primal sketch, specifies the locations of changes in light intensity level in the grey level image. The components (or "tokens") of the raw primal sketch represent specific items -- oriented edge segments, bars (parallel edges) and their terminations -- in the variation of light intensity across the grey level image. The third representation of the image, the full primal sketch, specifies the spatial organization of the image

[5]. Fodor (1983) extends this idea to an entire class of what he calls "input systems", which includes language comprehension.

implicit in the raw primal sketch. The purpose of the full primal sketch is to locate boundaries between regions of the image that reflect discontinuities in physical surfaces.

According to Marr, the next stage of processing generates a viewer-centred representation of the distance and orientation of visible surfaces, which Marr calls the 2 1/2-D sketch. The idea here is that the results of a number of independent computations operating on the primal sketch combine to produce a single representation of distal surfaces. In different ways most of these computations exploit the changes in spatial relations between tokens of the primal sketch that result from changes in viewer position.

In the last stage of processing, the 2 1/2-D sketch is used to generate a complete object-centred representation of objects and surfaces in the physical world. The basis of the theory here is a theorem according to which a particular representation of three-dimensional shapes called a generalized cone can, given certain assumptions about physical surfaces, be derived from surface discontinuities represented in the 2 1/2-D sketch.

In Marr's presentation of these theories, the computational level of description is restricted to the functions and representations that comprise this sequence. In these theories, the purpose of the computational level is to demonstrate that the computation of a particular function in the sequence -- for example, calculation of the three-dimensional shape of an object from a series of two-dimensional retinal projections in the generation of the 2 1/2-D sketch -- yields a true representation of the stimulus from available information; this is coupled with experimental evidence that this solution is in fact used in human vision. The algorithmic level of description in these theories is a set of formal schemes and algorithms that describe how each function in the computational-level sequence is realized by a set of symbolic transformations. Each distinct symbolic realization of the computational-level sequence of functions will perform differently for different arguments and under different circumstances. The purpose of the algorithmic level is to demonstrate that a particular formal scheme and algorithmic realization of the symbolic transformation function matches the performance of the visual system. The notion of "performance" at the algorithmic level includes such things as chronometric studies and measures of the accuracy of output under degradation of data. At the physical level of vision theory it is shown how a symbolic transformation is implemented in neural processes.

Hence, in his empirical theories the distinction between computational and algorithmic levels is "absolute"; i.e., whether a function in the theories is part of a computational-level description or part of an algorithmic-level description is not relative to a context of enquiry. In the theory of vision there is a distinguished set of functions that comprises a computational level of description of visual processes, and another distinguished set of functions that constitutes an algorithmic level of description of those same processes.

There are two grounds for identifying a particular level of decomposition as comprising a distinguished computational level of theory:

(1) As an explanation of the dependence of successful perception on certain global properties, or "natural constraints", of the physical environment, the computational level of description has explanatory significance not shared by decompositions at other levels of analysis.

(2) The explanatory role of the computational level dictates that the character of the computational level is determined by the relations in the stimulus formed by the natural constraints, and not by the steps in the behaviour of the system. Thus the computational level forms an autonomous level of theory.

In Section 3 I look in detail at these two points.

3. Decomposition at the computational level.

I claimed at the outset that Gibson's theory of perception is an important antecedent to Marr's approach to vision. To establish the autonomy of the computational level, we can begin by looking at

Gibson's influence on Marr's theory. The basis of Gibson's account is that spatial information about the world is obtained from what he refers to as "higher-order variables". The point of introducing the notion of higher-order variables is an assertion that relations in the stimulus upon which perceptual processes are based are not relations between first-order physical magnitudes such as distance or light intensity, but rates of change in those magnitudes. The simplest example of this idea is the relation between the slope of a distal surface and what Gibson calls "texture gradients" in the image.

As is familiar from perspective line drawing, constant increments in distance on a surface in front of the viewer are correlated with constant decrease in distance between the corresponding points in the image. And if the slope of the surface with respect to the viewer is altered, this is reflected in a change in the variation between distances in the image. Hence if there are features of the ground surface occurring at constant spatial distances from one another then the slope of the surface with respect to the viewer is a function of the rate of variation in distances between these features in the image. But many surfaces have a regular texture with elements occurring at constant distances so that the slope of surfaces facing the viewer can be obtained as a direct function of the rate of change in distances between texture elements in the image, which

Gibson calls texture gradient. The slope of surfaces in the distal scene is thus a function of a single higher-order variable in the image. By contrast, if we attempt to determine distances between single points in space from the distances between their counterparts in a static image, we face the problem that distances between features in the image are functions of both the distances between their correlates in the world and their distances from the viewer. Hence one of these must be known before the other can be determined.

While determination of surface slope from texture gradient depends on variation in relative distance across the image, Gibson argues that information can be obtained in a similar manner from variation in relative distance in the image over time due to changes in the position of the viewer. For example, he suggests that the three-dimensional shape of an object might be recovered from the variation in the shape of its two-dimensional projection in the image as the object moves with respect to the viewer, an idea fully developed in Ullman's shape-from-motion theorem. Deriving three-dimensional configurations from variation in their two-dimensional projections over time avoids the underdetermination of three-dimensional geometry by momentary two-dimensional stimulation.

According to Marr, Gibson's recognition of the importance of relations between higher-order physical magnitudes in the normal environment is his fundamental contribution to perception theory.

> Gibson's important contribution was to take the debate away from the philosophical considerations of sense-data and the affective qualities of sensation and to note instead that the important thing about the senses is that they are channels for perception of the real world outside or, in the case of vision, of the visible surfaces.
>
> (1982: p.29)

But what precisely is the significance of Gibson's work? It cannot be the recognition of higher-order variables per se. Ohm and Helmholtz developed the role of Fourier analysis in explanation of tonal perception, an idea that incorporates whatever can be meant by variables of higher-order. Rather Gibson's contribution lies in his insight that relations between rates of change in certain detectable magnitudes in the stimulus are constrained by the structure of the physical environment.[6] Higher-order variables in the stimulus can be exploited in the manner Gibson describes only if there are constant, or regular, spatial relations over time or distance between identifiable features of the surrounding surfaces. For example, the determination of surface slope from a texture gradient depends on the fact that natural surfaces have an identifiable texture with elements at constant distances. And as Ullman's shape-from-motion theorem makes explicit,

---

[6]. In Appendix 1 I discuss Gibson's theory of direct perception in more detail.

the recovery of shape from variation in retinal projections depends on the rigidity of physical objects. Hence the extraction of information in the manner suggested by Gibson depends on the physical structure of the environment. Thus the importance of "higher-order variables" for perception theory is that perceptual response is seen to exploit the existence of specific global properties of the normal environment. As a result, perception theory is extended from physiological optics to the study of properties of objects and surfaces in the world. Marr refers to these global properties of the physical world upon which early perceptual processes are based, as natural constraints.

By exploiting these constraints on relations between stimulus magnitudes, significant information about the distal scene can be obtained by highly modular processes. Natural constraints are global properties of the normal environment; so computations that deploy them can operate reliably with limited access to information about the current stimulus. For example, a process that derives surface slope from texture gradient need have access only to the value of a single variable in the stimulus. Similarly, the computational behaviour of a process that computes Ullman's shape-from-motion function is completely determined by the variation in the spatial relations between four points in the image. According to Marr's theory, early vision consists of a collection of such modular processes, each of which computes a single functional relation between magnitudes in the stimulus that is a consequence of natural constraints in the physical structure of the environment.

Let us consider the form of a computational-level description of such a process. The description must do three things: (1) It must provide a canonical description of the computation performed by the process; (2) it must describe the conditions under which such a function yields a true representation of a feature of the stimulus from that of another feature; (3) it must describe the global properties of the physical composition of the environment that makes these conditions likely to hold. The last two tasks confer a unique role on the description: The computational-level description of the process must relate the behaviour of the system to the structure of the physical world. To see this it is best to look at how such a description constitutes an autonomous level of theory.

A description of a modular process of the sort under discussion that accomplishes the three tasks just listed is a computational-level description of a particular class of computing mechanisms. It specifies a canonical encoding function for representing the values of physical magnitudes in the stimulus, and it demonstrates that under that encoding a canonical symbolic transformation map generates a true representation of one set of stimulus magnitudes from a true representation of another set in an environment that exhibits specific regularities.

Consider one such mechanism M_1, in isolation from other computational mechanisms of the system. Suppose that M_1 computes a function F_1 from the values of one set A_1 of stimulus magnitudes to the values of another such set A_2. An assertion that such a mechanism is part of the visual system entails the claim that the values of A_1 are "available" to the system; i.e., the system has somehow determined the values of A_1 as input to M_1. This in turn entails that there is some process of the system that produces a true representation of A_1. One possible explanation of this claim is that the values of A_1 are what I will call directly available to the system -- by this I mean that the explanation of how the system produces a true representation of A_1 involves only descriptions of the properties of A_1 itself, together with descriptions of the neurological processes of the system. When a system represents a magnitude of the stimulus in this way, I will say that the system directly detects that magnitude.

But in most of the cases described in Marr's theories the explanation of how A_1 is made available to M_1 is that the values of A_1 are obtained by computing another function F_0, where F_0 describes a relation between A_1 and the values of a third set A_0 of stimulus magnitudes which holds in virtue of a distinct set of natural constraints.[7] In this case we explain how a true representation of A_1 is obtained by giving a computational-level description of a second mechanism M_0 that computes F_0.

By coupling the two mechanisms so that the output of M_0 is the input to M_1, we have a description of a "composite mechanism" $M_{<1,0>}$ that computes values of M_2 from values of M_0. By describing $M_{<1,0>}$ as a composite of the two simpler mechanisms it is characterized as performing two distinct operations. But in this description, we have been providing only a computational-level theory of the composite mechanism; what we have is a description of a virtual machine which demonstrates only that the requisite information is available by mechanical operations. To explain the mechanical operations by which the information is in fact extracted we have to supply algorithmic-level and physical-level descriptions of the composite mechanism $M_{<1,0>}$.

Now consider two possible algorithmic-level descriptions of $M_{<1,0>}$:

(1) On one such description the algorithm includes the sequence $<F_0, F_1>$. This means that there is an encoding function that encodes the possible possible values of A_0, A_1 and A_2; and there are two symbolic transformation functions T_0 and T_1 such that under the encodings T_0

---

[7]. Eventually of course this regress must terminate at some set of stimulus magnitudes that is directly available.

takes representations of A_0 to representations of A_1, and T_1 takes representations of A_1 to representations of A_2. So this mechanism performs two distinct steps in the computation of the values of A_2. (2) On a second algorithmic-level description the encoding function is a subfunction of the first that assigns symbolic representations only to values of A_0 and A_2. There is then a single symbolic transformation function $T_{<0,1>}$ extensionally equivalent to the sequence $<T_0, T_1>$. In this case there is a single computational step in the generation of the final output. But this second description is an algorithmic level description of the same computational-level mechanism $M_{<0,1>}$ because it yields the same representation of A_2 for a given representation of A_0, so that it is truth-preserving just in case the first one is.

If we suppose that one of the two algorithmic-level descriptions is true, what decides which one it is? Clearly this is a matter of which description is realized in the physical-level description of the mechanism. According to the first algorithmic-level description there is an interpretation function I that maps physical states of the mechanism to the symbolic representations of A_0, A_1 and A_2. And there are on this description two physical-state transition functions P_0 and P_1 such that under I the mechanism computes T_0 and T_1. According to the second algorithmic-level description the interpretation function I′ is a subfunction of I that maps physical states of the mechanism to the symbolic representations of A_0 and A_1. And on this description there is a single physical-state transition function such that under I' the mechanism computes $T_{<0,1>}$.

The point I want to stress is that the two algorithmic-level descriptions have the <u>same</u> computational-level description. In particular, the computational-level description must decompose the process into a description of <u>two</u> virtual mechanisms <u>even if the algorithmic-level decomposition consists only of a single symbolic operation $T_{<0,1>}$</u>. This is because the computational-level description must relate the truth-value of the output to the natural constraints on which it rests. The role of the natural constraints in the explanation of the operation is to constrain relations between A_0, A_1 and A_2 so that they conform to the functions F_0 and F_1. Hence describing the relation between the natural constraints and the veridicality of the operation is possible only by describing the relation between the initial and final representations as a computation of the two functions F_0 and F_1.[8]

The same reasoning applies to any set of algorithmic-level descriptions of mechanisms whose initial and final states exhibit the same semantic relations under the same natural constraints. <u>Hence the character of the computational-level description is dictated by the relations between physical magnitudes upon which the veridicality of</u>

---

[8]. This point is illustrated in Appendix 2 by a discussion of Marr and Poggio's (1979) theory of stereoscopic vision.

the computation rests, and not by the steps involved in the mechanical production of the final output.

So the computational-level decomposition of visual processes that relates the veridicality of the operations to the natural constraints upon which that veridicality rests is autonomous with respect to the descriptions which characterize the mechanical realization of the input-output pairs. For this reason Marr is justified in identifying a distinguished computational-level description of the visual system. The sequence of functions and representations that we reviewed in Section 2 comprise precisely such a decomposition of visual processes. As it happens, the algorithmic-level descriptions provided by Marr are such that each function in the computational-level sequence is computed in the algorithmic-level decomposition of the system. But as we have seen, this issue is decided by facts other than those that determine the character of the computational-level description.

4. Summary.

Let us briefly review the results of this chapter. We saw the role of Marr's computational level of description is to relate the computational behaviour of a mechanism to its intended purpose. This requires a canonical description of the function computed by the mechanism, together with a demonstration that computation of that function yields information required from information that is available. The algorithmic level of description then specifies the symbolic transformation functions by which this computation is carried out by the system, while the physical level describes how these transformations are computed by physical-state transitions of the system.

It was noted that in its general application, the distinction between computational and algorithmic levels of description is relativized to a context of enquiry. But in its application to explanations of early vision, there is a distinguished computational-level description that explains how true representations of stimulus magnitudes are obtained through exploitation of natural constraints in the normal environment. In the Section 3 we saw that an absolute distinction between computational and algorithmic levels in early vision is warranted by the fact that the character of the computational-level decomposition is autonomous with respect the steps in the algorithmic level of description. In the chapters that follow I apply these facts about the character of the computational level to philosophical assertions about the nature of computational theories.

CHAPTER TWO

SUPERVENIENCE AND COMPUTATIONAL EXPLANATION

0. Introduction

In Chapter 1 we saw that a central aspect of the computational theory of vision developed by Marr is the use made of contingent regularities, or "natural constraints", in the physical environment to explain how the visual system determines the shape and location of objects in the world on the basis of the spatial organization of the retinal image. In this chapter I am concerned with a recent discussion of Burge (1986) concerning the implications of this feature of Marr's theory for understanding psychological explanation.

Burge claims that Marr's use of natural constraints shows the theory to be "nonindividualistic". By this he means that the individuation of psychological states within the theory depends essentially on the objects and conditions of the world external to the subject. In various papers Burge has given a number of versions of individualism. Here I am concerned only with one version, namely, the rule that the "representational content" of cognitive states "supervenes" on neural states. To motivate this issue, let me first introduce some terms.

Let us call the representational states of a cognitive system the set of neural states partitioned by their "representational content". It is very hard to say what representational content is in general; but in the restricted context of early vision we can say that the representational content of a state is the information it is held to carry under a computational description. Then the rule that representational states supervene on neural states is the requirement that there are no differences among representational states for which there is not a corresponding difference among neural states. So supervenience is a constraint on interpretation functions to the effect that the same neural state cannot be mapped to more than one representational state. Against this view of psychological theories, Burge argues that the role of natural constraints within Marr's theoretical framework has the consequence that the individuation of representational states changes as we (counterfactually) vary the description of the surrounding environment, while the subject's physical history remains fixed.

In this chapter I argue Burge is correct in his claim that Marr's theory is nonindividualistic. That is, in some contexts Marr's explanatory framework justifies assignment of different interpretations to systems that have identical physical descriptions but operate in different environments. However, to avoid trivializing computational explanations it is important to recognize that there are restrictions on the ways in which supervenience can be relaxed, and in this respect Burge's account is misleadingly incomplete. Here it is useful to compare the issue of individualism with the question whether Marr's theory is compatible with methodological solipsism as a research strategy. In Putnam's original formulation, methodological solipsism is the rule that "no psychological state, properly so-called, [should presuppose] the existence of any individual other than the subject to whom that state is ascribed" (Putnam 1975, p.220). In some discussions methodological solipsism is similar in content to individualism. But in Fodor's 1980 paper it assumes a very different form. Fodor maintains that methodological solipsism, realized as the assumption that semantic properties of psychological states are irrelevant in explanations of behaviour, defines the approach to theory construction of computational psychology. I show that Marr's theory is committed to methodological solipsism under Fodor's formulation, and that this fact provides important constraints on relaxation of supervenience. But Fodor's description of the role of methodological solipsism in theory construction suffers from a confusion of issues which has generated unnecessary resistance to his thesis. This confusion is revealed in Kitcher's (1988) contention, argued on grounds similar to those advanced by Burge vis a vis individualism, that Marr's theory violates methodological solipsism. In the concluding section I argue that Kitcher's point is well-motivated but incorrect as it is stated.

1. The "Argument from Success".

Let us first consider the claim that Marr's theory is nonindividualistic. Burge's argument for this claim rests on two assertions about Marr's theories:

(A) The theories are designed to explain our success at certain visual tasks, so that representational states are specified in a way that will account for the veridicality of perception. In particular, this means that the content of the visual representations specified by the theory must be true in the normal case. Thus the content (and hence the individuation) of representational states is determined by their normal causal antecedents.

(B) The explanations the theory provides assume that the visual system has evolved so as to exploit natural constraints in the normal environment; our abilities are thus taken to require that global regularities in the physical world ensure that early computational processes are truth-preserving. In this way, the causal links between states of the world and representational states of the system depend on contingent facts about the normal environment.

Burge argues from these two premises to the nonindividualism of the theory by constructing a thought experiment similar to Putnam's Twin-Earth story. We can imagine two individuals who have identical physical histories but who live successfully in environments that differ in the regularities that underlie each individual's visual abilities. The point of the story is that unless the representational states of the subjects are individuated differently in the two cases, one of them would have representations that are regularly false contrary to (A) above. Hence we must assign different representations to their perceptual states despite their physical similarity. Burge concludes that the dependence of Marr's theory on the specification of natural constraints that underlie the relations between visual representations and the world requires the ascription of type-distinct representational states to the two subjects, <u>despite their physical type-identity</u>:

> The methods of individuation and explanation are governed by the assumption that the subject has adapted to his or her environment sufficiently to obtain veridical information from it under certain normal conditions. If the properties and relations that normally caused visual impressions were regularly different from what they are, the individual would obtain different information and have visual experiences with different intentional content. (Burge 1986, p.35)

Hence, it is argued, on Marr's approach to vision representational states of a subject do not supervene on neural states. Let's look first at what is right about Burge's argument.

We have seen in the preceding chapters that both of (A) and (B) are true. Take (B) first. On Marr's view, the explanation we give of visual perception depends crucially on the nature of the environment. This is an important consequence of Marr's argument for a computational level of description. The problem at the computational level is to understand how veridical representations of the world can be obtained from light intensity values in the image. As we have seen, Marr's solution to this problem is to find features of the stimulus, light intensity gradients or patterns in the geometrical structure at different scales for example, that carry reliable information about the world, and that can be used in determining the final representation. A crucial factor in Marr's approach to the problem is that only a study of the contingent regularities in the actual environment will reveal which image properties are reliable carriers of useful information.

(A) is equally important for understanding the theory. The computational level of theory is formulated to explain the veridicality of perception. At <u>that</u> level of the theory the representations postulated are those required to show how the success of visual processes is achieved. Specifically, <u>the truth-values of intermediary representations in the computational-level sequence have an important</u>

role in explaining the veridicality of the final representation -- if they are illusory in the normal environment, so is the final representation.

However, Burge's discussion suggests a different, and mistaken, reading of (A). It is a matter of common observation that perceptions are generally veridical in the normal environment. We can interpret (A) as simply the assertion that the role of perception theory is to specify causal connections between representations and properties of the world that account for this observation. On this view we do not require that the veridicality of any one representation depends on the veridicality of others. Rather the veridicality of each is seen to be a consequence of a distinct correlation between it and some feature of the stimulus in such a way that the explanations of these correlations are independent of each other. This reading is similar to Gibson's account. According to Gibson, theory construction is a matter of discovering, for each property P of the distal scene that we are able to successfully perceive, a set of variables in the ambient light that is reliably correlated with P.

But Gibson's view stems from his mistaken emphasis on higher-order variables. By pointing to such variables Gibson argues against the belief that perception involves information-processing, i.e., against the belief that vision is a product of operations on representations. However, as we have seen, the significance of environmental constraints on higher-order variables is not that it eliminates visual processing; rather it shows how such processing can be accomplished by modular units. A crucial feature of Marr's theories is the fact that the computational-level functions and representations form ordered sets such that the information carried by one representation provides input to later operations. The existence of natural constraints allows for the operations in the sequence to have access to very limited and specific information, and yet collectively to generate a complete representation of the distal scene. So the significance of (A) lies in the fact that the truth-value of each representation in the computational-level sequence contributes to the explanation of the truth-value of the final representation.

So Burge is right in pointing out that we cannot construct a theory of vision like Marr's if we restrict our attention to the physiological states of the subject. Understanding perception requires a knowledge of how the patterns of retinal stimulation are causally linked to appropriate features of the world. According to Marr's theory, states of the system carry information about conditions in the world because, as long as the natural constraints ensure the causal links, they are regularly correlated with such objective conditions. Thus the content of representations postulated to explain perception depends on the structure of the physical environment, and this is justified by the presumption that we have evolved so as to exploit natural constraints in our surroundings that connect the structure of light intensities with shapes and surfaces in the world.

The question, then, is whether the denial of supervenience is a consequence of (A) and (B) in the manner suggested by the thought-experiment. Against this conclusion, Fodor (1987, Chapter 2) maintains that the possibility of mental causation depends on supervenience --if Burge is right we will be unable to explain how cognitive processes are realized in neural structure. Let us look first at Fodor's argument.

2. Assessment of the Argument from Success.

The issue of individualism is often discussed as the general question whether certain relational properties of mental states are relevant to psychological taxonomies: Is the type-identity of the psychological states of a subject affected by relations between the subject and the external world? But according to Fodor, the real point in the defence of individualism is not the relevance of relational properties; rather it is the contention that psychological taxonomies should not include distinctions that are not causally relevant. A taxonomy is individualistic as Fodor uses the term if it "distinguishes between things insofar as they have different causal properties, and ...

groups things together insofar as they have the same causal properties." (1987, p.34) A causal property of an object is a property in virtue of which the object is subsumed under a causal generalization. He contends -- correctly, it would seem -- that the requirement that theories are individualistic in this sense is constitutive of scientific taxonomies generally.

Fodor argues that changes in the environmental surroundings of a subject that do not affect the subject's neural states can have no relevance to the causal generalizations into which those states can enter. In particular, distinctions among representational states that do not correspond to differences among neural states can in no way affect the behaviour of the system.[1] Thus such taxonomies remove the basis for descriptions of mental causation of behaviour. So any taxonomy that does not preserve the supervenience of representational states draws distinctions that reflect no differences in the causal properties of representational states. Since this violates the just noted requirement on scientific theories, supervenience is a first principle of psychological explanation.

Fodor is certainly right that psychological explanations of behaviour cannot appeal to differences in the representational states of a subject that do not have a corresponding difference in neural states.

[1]. Unless you individuate behaviour the same way, which leads to absurd results.

But that constraint is not inconsistent with nonsupervenience in the thought-experiments. The generalizations that underlie computational explanations are not causal generalizations -- at least not in the sense that the taxonomies induced by computational descriptions are based on sameness of causal properties. Rather the generalizations captured by computational descriptions are those that assert an equivalence of different physical systems under a single computational description. According to Fodor, to deny supervenience is to construct a taxonomy that is not based on causal properties. But this is incorrect. The issue is whether there are legitimate explanatory reasons for assigning different computational descriptions to the same physical mechanism. Certainly, a physical system may fall under more than one computational description -- trivially so, since under some description every system computes every function. The thought-experiments are intended to show that there are good explanatory grounds for altering the computational description of a system under different environmental conditions.

Burge's argument from the thought experiments is sound. We can reformulate the argument according to the characterization of computational theories laid out in chapter 1. Let us suppose that as a consequence of some set of natural constraints in the normal environment, the value of a physical magnitude M_1 in the stimulus is a function F of another stimulus magnitude M_0. Let us also suppose that according to a theory Γ, the visual system determines the value of M_1 by computing this relation between M_0 and M_1. If it is a complete theory, Γ describes a physical-state transition function P over computational states of the system; and it provides an interpretation function I, where I maps states in the domain of P to M_0 and states in the range of P to M_1.[2] Burge's thought experiment describes a situation in which there is a <u>distinct theory</u> Γ′ for the system that describes the same physical-state transition function P, but where the interpretation function I′ maps states in the range of P to values of a magnitude M'_1 distinct from M_1. Then Burge's argument is that Γ′ will be the preferred theory of the system if the following two conditions are met: (1) the system is in an environment where, given some new set of natural constraints, M'_1 is a function F′ of M_0; (2) the system computes F′ under P and I′.

In fact, since the computation of M'_1 will most likely be a stage in a sequence of computations that generates a composite representation, we must suppose that a similar relation holds between M'_1 and other magnitudes in the stimulus, and that there are interpretation functions

[2]. In most cases the values of M_0 and M_1 will be represented by symbolic expressions under an encoding function, and F will be computed by carrying out a symbolic transformation function. But these complications do not alter the point, so they can be ignored here.

that describe the state transitions of the system as computing these relations. But these conditions do not appear impossible, and Burge's conclusion is consistent with Marr's explanatory framework.

So individualism is false with respect Marr's theories. However, I believe that the significance of this lies in the emphasis it places on Burge's two premises (A) and (B), which are important. There is a danger in identifying nonindividualism as the most significant consequence of (A) and (B) in the way Burge does. Nonindividualism tells us that in certain contexts it is legitimate to assign distinct interpretations to the same neural processes. Burge's account of Marr's theory says that such assignments are warranted in explanations of early vision in order to preserve the overall veridicality of the subject's representations. But clearly there must be constraints placed on the assignment of interpretations in computational theories to ensure the explanatory power of the theory. It is precisely the specification of these constraints that must be spelled out to determine the nature of computational explanation. For example, the trivialization problem that attaches to Gibson's theory is that, although he successfully identifies causal links between stimulus features that ensure the veridicality of perception, he does not specify the computations by which these causal links are exploited in the mechanical generation of representations. Fodor's concerns about the mechanical production of behaviour by representational states still needs a solution. In the next section I will claim that methodological solipsism is an important constraint to ensure the explanatory power of computational explanation, and in particular that it sets important limits on the relaxation of supervenience.

3. Supervenience and methodological solipsism.

In the introduction I claimed that in Fodor's (1980) discussion methodological solipsism is different in content from Putnam's original formulation. In his statement of the principle Fodor presents methodological solipsism as a claim similar to Putnam's version, and thus to some extent similar also to Burge's individualism. But Fodor's application of methodological solipsism concerns different issues than those that which motivates both Burge's and Putnam's discussions.

In his statement of methodological solipsism Fodor says that the principle is represented in computational psychology by what he calls the "formality condition". The formality condition says that "two thoughts can be distinct in content only if they can be identified with relations to formally distinct representations". (1980: p.227) By "formal properties" Fodor intends any nonsemantic properties of psychological states, where semantic properties include truth, reference and representational content. In theory of early vision, where the difficult problems in specifying the representational content of propositional attitudes do not arise, we can take the representational content of a computational state to be a function from physical environments and computational states to truth-values. So the formality condition says that the type-identity of representational states is not affected by changes in the environment of a system that do not affect the physical or symbolic character of computational states.[3] This much appears as a version of Burge's individualism.

Fodor's version of methodological solipsism breaks with individualism in its application as a constraint on computational explanations of behaviour. Here is Fodor's statement of how methodological solipsism functions in computational psychology.

> I'm saying, in effect, that the formality condition, viewed in this context, is tantamount to a sort of methodological solipsism. If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide. Hence, they have no access to the semantic properties of such representations, including the property of being true, of having referents, or, indeed, the property of being representations of the environment. (1980: p.231. Italics in the original)

---

[3]. For Fodor appeal to symbolic transformations is constitutive of computational explanations. So on his view, formal properties are properties of symbol sets. I think, however, that this is not true of very early visual computations such as edge detection. So I will include both physical and symbolic states here.

Notice that in this statement methodological solipsism is an assertion about the kinds of properties to which mental processes have access. Fodor's methodological solipsism is concerned with placing constraints on what can enter into descriptions of domains over which mental processes are defined. The point of methodological solipsism, as Fodor formulates it, is that any distinctions between mental states that affect the behaviour of the system must correspond to distinctions between physical properties of the system.

By contrast, individualism is not a restriction on the properties of mental states that influence behaviour. Rather, it is a constraint on psychological explanation generally, viz., psychological explanations must not appeal to distinctions among representational states to which the processes of the system have no access. There is no conflict between methodological solipsism and a denial of individualism, for they are concerned with different things. Methodological solipsism says that the processes of the system have no access to semantic properties of mental states, so that these properties cannot affect the behaviour of the system. By denying individualism we allow only that semantic properties of computational states can be appealed to in psychological explanations for some purpose, where it is left open how this appeal can appear within the theory. So an appeal to semantic properties that violates individualism is not inconsistent with methodological solipsism as long as it does not appear in the description of the domains over which mental processes are defined.

The appearance of conflict between the two doctrines arises from Fodor's remarks later in his paper which describe methodological solipsism as restricting the domain of what is relevant to psychological explanations. Thus, for example, Fodor remarks that "there can't be a psychology of knowledge." (1980: p.228) There is, then, a question why he broadens the scope of methodological solipsism in this way. I think that the inference to the conclusion that methodological solipsism is a restriction on psychological explanation generally has the following form:

> The business of psychology is to provide explanations of the mental causes of behaviour. According to methodological solipsism, the mental processes that generate behaviour have no access to semantic properties of representations. Therefore semantic properties are irrelevant to psychology.

But we have seen that semantic properties <u>are</u> relevant to psychological explanation, and this is the thrust of Burge's argument. Indeed, given the role of the computational level of description isolated in Chapter 1, formulating a "psychology of knowledge" is the first part of theory construction. On the presumption that we have evolved to exploit the effects of natural constraints, the character of cognitive processes is determined in part by semantic properties of representational states. However, this fact doesn't entail the conclusion that methodological solipsism is false. Rather it entails that methodological solipsism is a constraint on <u>how</u> semantic properties can enter into psychological explanations. In this form methodological solipsism is both true and important.

So we need to specify how semantic properties can enter into computational explanations of perceptual abilities. I think that the answer now is quite clear. <u>Semantic properties can appear in explanations of perception as part of the justification of the choice of a computational theory as characterizing a particular mechanism in a particular environment</u>. This statement explains the basis of nonindividualism. Burge's thought-experiment is intended to provide an example of a case in which computational theories of two systems in different environments may differ only in the interpretation function that takes computational states of the system to properties of the environment. <u>But by methodological solipsism semantic properties cannot appear within a computational theory as part of the individuation of representational states over which the computations are defined</u>.

Fodor reaches much the same conclusion following his discussion of Burge. He points out that, in justifying the choice of a particular theory, the "psychologist's taxonomic apparatus is, often enough, nonsolipsistic". He concludes,

> These sorts of explanations square with individualism, because the relational facts they advert to affect the causal powers of mental states; indeed, they affect their very existence. But naturally, explanations of this sort -- for that matter, all teleological explanations -- are ipso facto nonsolipsistic. (Fodor 1987, p.44)

But there are two problems with making the point this way. (1) Fodor's account rests on his appeal to causal properties, and this is not an appropriate way to frame the explanatory form of computational theories. (2) If we take this approach, theories like Marr's will yield two taxonomies of representational states depending on our explanatory goals. If we take causal properties of these states to be specified by their role in the production of behaviour, then since the behaviour of the visual system is unaffected by the semantic values of its representations, semantic properties will not be included. Thus representations will not be individuated by semantic properties if the taxonomy is designed solely to account for the behaviour of the system. But as Burge has shown, if our purpose is to explain why this behaviour is successful we will have to specify the content of representations in order to account for their veridicality. Thus we will pull the theory in two ways, depending on whether or not we count the reliability of perception to be part of what it is designed to explain.

But it is quite unnecessary to state things in this way. While we want to allow for the theoretical utility of semantic properties, we do not want to go so far as to give truth values a role in explanations of behaviour. But this is not a problem at all: The requirement that truth values not enter explanations of behaviour is unaffected by what for Burge is a nonindividualistic taxonomy of states. We can allow that representational states do not supervene on physical states, in the sense that computational theories of the same physical system may differ only in their assignment of representational content to computational states. We do not thereby sacrifice the explanatory power of computational explanations as long as we require that within a theory there are no differences in representational content -- and accordingly no differences in truth values -- that do not correspond to a difference between physical states of the system. Fodor is quite right in conceding that "teleological" arguments that justify the selection of one computational description over another are nonsolipsistic. But the claim that teleological considerations affect how we describe the system, including the individuation of mental states, does not preclude the possibility that computational descriptions of behaviour are entirely solipsistic.

In a discussion of Marr's theory, Kitcher gives the following argument against the conclusions of this section.

> Most directly, if Marr is right, then a theory of vision must incorporate information about the environment, both in describing the representations produced by the system and in describing the constraints that it uses to disambiguate information in the grey-level array. Thus, Marr's project violates Fodor's canon of Methodological Solipsism, because it does not confine itself to syntactic or formal features of internal representations; rather it makes essential reference to factors beyond the subject's skin in characterizing psychological states. (Kitcher 1988, pp. 13-4)

Kitcher's argument is clearly directed at Fodor's extension of methodological solipsism to a denial of the relevance of semantic properties in psychological theories generally. Notice that, as Kitcher understands it, methodological solipsism is the assertion that psychology must restrict itself to the "syntactic or formal features" of representations. On her reading, methodological solipsism denies a place in psychological theories for a computational level of description; thus Kitcher maintains that methodological solipsism is inconsistent with Marr's explanatory framework for computational theories.

Kitcher is correct in her claim that Marr's theory is shaped by facts about the environment. But as we have just seen, the explanation the theory gives of the computational behaviour of the system appeals only to nonsemantic properties of representations. Like Burge, Kitcher fails to distinguish the use of environmental regularities to explain the success of visual mechanisms from the description of those mechanisms in the theory itself. Marr's commitment to methodological solipsism is manifest in his insistence on the latter.

4. Summary.

In Section 1 we saw that Burge's argument against individualism identifies two important features of Marr's theory: (A) The theory is designed to explain the success of visual perception so that representational content must be assigned so that representations are true in the normal environment; (B) according to Marr, the veridicality of visual perception depends on contingent facts about the physical world. In Section 2 I argue that Burge's argument is sound; that is, for the reasons he cites, there are explanatory reasons for assigning different computational interpretations to systems in different physical environments that are physiologically identical. Fodor's argument against this is based on an incorrect understanding of the nature of computational theories. According to Fodor, computational theories specify causal properties of representational states; but the correct point is that the explanatory generalizations of computational theories assert that different physical systems are equivalent under a computational description. However, there is a danger in placing undue emphasis on the issue of individualism. And here it is important to see that methodological solipsism captures a crucial constraint on computational theories: While there are distinct computational descriptions of the same physical system that differ only in the interpretations assigned to its computational states, within a theory there can be no differences in content for which there is not a corresponding difference in neural states. Fodor (1980) generates unnecessary opposition to this point by extending methodological solipsism to a restriction on the explanatory goals of computational psychology. It this is this restriction that Kitcher's criticism attacks, and not the real point of methodological solipsism.

CHAPTER THREE

AUTONOMY AND REDUCTION OF PSYCHOLOGICAL THEORIES

0. Introduction.

In Chapter 1 I argued that there is a specific sense in which the computational level of theory constitutes an autonomous description of the behaviour of the visual system: The computational-level decomposition of the relation between the image and the final representation is determined by relations among properties of the stimulus, and is unaffected by the steps carried out in the computation of that relation by the mechanisms of the system. And in Chapter 2 we saw that, while supervenience of representational properties on neural properties is an essential feature of computational theories, there are explanatory grounds for assigning distinct computational-level theories to systems that have identical neurophysiological descriptions. In this section I will use these points to argue that <u>reductionism</u>, i.e., the claim that psychology is reducible to neurophysiology, is false. Although the details are not clear, in the broadest terms we will say that a special science is reducible to a more basic science just in case (1) the properties of the special science are identical to, or correspond in some law-like way, to properties of the basic science, and (2) the principles of the special science are consequences of the principles of the basic science. I will refer to (1) as physicalism.

My arguments will be directed primarily at recent discussions of reductionism by Churchland (1986) and Kitcher (1988). Before turning to these discussions I address in Section 1 a common argument against reductionism based on a denial of physicalism. This argument, I believe, is not sound. The point I wish to highlight in Section 1 is that reductionism ought not be treated as a metaphysical issue, but rather as one of explanatory method. In sections 2 and 3 I look at an argument by Kitcher against reductionism, based on the explanatory role of Marr's computational level of theory, and at a contrary conclusion of Churchland's. Kitcher's argument, though incompletely stated, is sound; Churchland's assertions are insufficient to establish reductionism, partly because Churchland sees antireductionism as based on a denial of physicalism. In the closing section of the chapter I extend these arguments to the claim of Kitcher and others that Marr's theory depends on the view that vision employs "optimal" strategies to solve perceptual problems. I argue that this claim is based on a misunderstanding of the autonomous character of the computational level.

1. Reduction and physicalism.

Functionalists argue from the multiple realizability of functional states that mental states do not correspond to neural states, and hence psychology is not reducible to neurology.[1] The most articulate version of this argument is by Fodor. According to Fodor, reduction implies that the following two conditions are met:

(1) For each property S of the special science there is a "bridge principle" of the form,

$$(x)\ (Sx <\!\!-\!\!-\!\!> Px),$$

where P is a property of the basic science. The biconditional must cover all possible realizations of S by properties of the basic science. In the most straightforward case, the properties will be identical, otherwise the biconditional is a contingent law.

(2) If $(x)\ (S_1 \dashrightarrow S_2)$ is a law of the special science, and

$$(x)\ (S_1 <\!\!-\!\!-\!\!> P_1),$$
$$(x)\ (S_2 <\!\!-\!\!-\!\!> P_2)$$

are bridge principles, then $(x)\ (P_1 \dashrightarrow P_2)$ is a law of the basic science.

Given this analysis, the argument against reduction is this: Mental states are functional states, determined by their causal relations to specified inputs and outputs and to other functional states. But the same functional state is realizable in very different neural structures. The set of neural structures that may realize a single mental state will be capturable in neurophysiological terms only by an indefinite disjunction of distinct properties; hence mental states will not correspond to any "natural kinds" of neurophysiological theory.

---

[1]. See Putnam (1967), Davidson (1970) and Fodor (1974).

Moreover, even if we were to possess a complete list of the neural properties with which a mental state co-occurs, such a list would never be sufficient to establish that the co-occurrences are lawlike. Thus physicalism is false, and hence reductionism is false also.

It is possible to extend functionalism, and hence also the argument just described, to embrace computationalism, i.e., the assertion that mental processes are computations. Recall from Chapter 1 that a mechanism M computes a function F just in case there is a physical-state transition map P:C --> C, where C is the set of states of M under the interpretation function I, such that for $c \in C$,

$$I(P(c)) = F(I(c)).$$

A formal representation of the state-transition diagram of a system yields a <u>functional</u> description of a class of mechanisms that are isomorphic with respect to their state-transitions diagrams -- the states are defined in terms of their relations to one another and to inputs and outputs. Thus computationalism can be seen to support the functionalist rejection of physicalism. It is an analysis of this sort that supports Fodor's view, discussed in Chapter 2, that mental states are representational states with causal properties.

However, there is an effective reply to the functionalist argument against physicalism. A closer look at scientific practice does not easily support the claim that a multiplicity of neural realizations of functional or computational states is inconsistent with identity of properties.[2] It is incorrect, so the reply goes, to suppose that physicalism demands a one-to-one mapping from mental properties to physical properties. Reductions between scientific theories are commonly specific to a domain of phenomena. For example, whereas opponents of physicalism cite the relation between thermodynamics and statistical mechanics as the best example of successful reduction, temperature in thermodynamics does not correspond to any single microscopic property of substances -- the description in statistical mechanics of the properties that correspond to temperature is an indefinite disjunction of precisely the sort described in the functionalist denial of physicalism. Churchland (1986) adds to this reply the claim that if the reduction of mental states is specific to humans then it is not likely that mental states will be correspond to very different neural states. Given our common evolutionary history, it is to reasonable to expect that any particular mental state will have a common set of neurological properties across individuals at some level of organization. Those that don't are likely to be dispensed with in favour of more successful explanatory kinds.

The defence of physicalism just described reveals a general flaw in the functionalist denial of the reducibility of psychology to

[2]. See Wilson (1985), Patricia Churchland (1986) and Paul Churchland (1984).

neurophysiology: Functionalism and computationalism are best advanced as theories of explanatory method rather than as metaphysical theories. In Chapter 2 I argued that, contrary to Fodor's analysis, computational theories are not causal generalizations, but rather assertions that there are good explanatory reasons for assigning a particular computational description to a physical system. Computational properties are not causal properties, and thus it is incorrect to identify computational states as explanatory kinds distinct from physical states; computational states are physical states under a computational description. It is Fodor's mistaken understanding of computationalism that supports his view that computationalism is inconsistent with physicalism. Functional properties, on the other hand, stand in causal relations by definition; but the argument that functional properties are physical properties is very strong. So there is evidence that a denial of reductionism on grounds of the falsity of physicalism is inconsistent with scientific practice, and the extension of the functionalist argument to computationalism rests on a mistaken analysis of computational theories. However, in the following sections I argue that reductionism is false on other grounds. In the next section I consider an argument of Kitcher's, based on Marr's explanatory framework, that neurophysiology is explanatorily incomplete. If this is correct then psychological principles are not consequences of neurophysiological principles.

2. Kitcher's argument from explanatory incompleteness.

According to Kitcher, psychology is not reducible to neurophysiology because of the explanatory incompleteness of neurophysiology. Kitcher agrees with Marr that neurophysiological research has a significant part in explanations of visual abilities. But on Kitcher's view, reductionists such as Churchland and Searle take the stronger position that only neurophysiology can provide scientific explanations of mental phenomena. Her discussion is thus directed at what she refers to as "the hegemony of neurophysiology". The argument that Kitcher offers against this form of reductionism assumes without argument that providing computational-level descriptions of visual processes is outside the scope of neurophysiological theory. It follows that, on her view, reductionism is inconsistent with the claim that a computational level of description is essential to explanations of vision. Thus, according to Kitcher, reductionism can be rejected if it can be shown that a computational level of description is needed to explain visual processes.

Kitcher's argument that a computational level of description is necessary makes two assertions:

(1) The computational level of description specifies the "information-processing tasks" performed by the system; i.e., the computational level describes the information about the stimulus that each computational process extracts. Models of the system not based on a computational-level description may mimic the behaviour of the system over a limited domain but will not succeed in explaining why the model is successful, or why it fails outside of its restricted domain.

(2) Determining the solutions to information-processing problems implemented by the system requires a description of the properties of the environment -- the natural constraints -- on the basis of which the system obtains information about the world. Here Kitcher appeals to the influence of Gibson's notion of stimulus information on Marr's approach to vision. Isolating natural constraints is not a neurophysiological task, but rather a part of the computational level of theory.

Kitcher is correct that algorithmic models of visual processes alone will not explain vision. As we saw in some detail in Chapter 1, we can explain the success of visual perception only by providing a computational-level description of relations among properties of the stimulus. But Kitcher offers no explanation of her view that a recognition of the role played by computational-level descriptions is inconsistent with the understanding of neurophysiological theory held by reductionists.

Some support for Kitcher's understanding of neurophysiological theory is found in her comparison of Marr's approach to vision with the feature-detector theory, a theory that was an important antecedent to Marr's work. The feature-detector strategy is based on the discovery that responses of single cells in the neuronal structure of the visual system signal the presence of patterns of light intensities in specific regions of the retinal image. Thus individual cells can be seen as performing complex tasks that extract specific information about the stimulus. From this organization in the neural structure of the system it is argued that (in some not fully specified sense) higher-level cells detect specific <u>features</u> in the image, and that recognition of objects in the distal scene is carried out by identifying collections of image features. According to Barlow (1972), the <u>methodological</u> import of the feature-detector approach is that much of perceptual psychology can be carried out directly by neurophysiological investigation: If perception is accomplished through selective response of individual cells to features in the image, modelling visual abilities can be accomplished by direct observation of the relations between features of the retinal image and the behaviour of single cells.

Kitcher's characterization of reductionism is similar to Marr's description of the limitations on the feature-detector program. Part of the failure of the feature-detector approach lies in the fact that interesting properties of physical objects do not correspond to collections of image features, contrary to the assumptions of the program. Marr's solution to this problem is the use of natural constraints in extracting information from the image in a succession of modular computations. But, on Marr's view, a <u>deeper</u> flaw in the feature-detector studies is a lack of attention to the information about the stimulus that an image configuration carries. In these theories, the term "feature" is applied indiscriminately to both the image and the distal stimulus, thus ignoring the question how features in the image relate to properties of the distal stimulus. Marr argues that the complexity of the relation between the image and the physical world is such that we can specify precisely the image patterns to which a cell is responsive only if we know what information about the distal stimulus such a pattern is interpreted as carrying. Hence we cannot extrapolate from recordings of cell behaviour to a computational theory of the system without a knowledge of how patterns in the retinal image correspond to properties of the physical world. The feature-detector program does not have a framework for investigating these relations, and so it fails to yield clear results.[3]

But an appeal to the shortcomings of the feature detector program is not sufficient to establish the general claim that computational-level descriptions lie outside the scope of neurophysiology. The view of neurophysiology that is presented by Churchland (1986) in particular is not constrained by the assumptions of the feature-detector theory. Churchland agrees that vision involves modular computational processes of the sort described by Marr; but she argues that computational theories of these processes are neurophysiological theories. In the next section I address Churchland's argument.

3. Reduction and autonomy.

According to Churchland, it is not true that neuroscience is restricted to constructing models of neural behaviour that instantiate computational theories constructed by psychologists. On her view, it is incorrect to suppose that there is a level of description of cognitive processes that is autonomous with respect to neuroscience. Her conclusion is based on several assertions:

---

[3]. In Appendix 3 I illustrate this point through a discussion of Marr and Hildreth's (1980) computational-level theory of the feature-detector studies.

(1) Neuroscience is concerned with specifying information-processing tasks performed by neural activity. She says,

> It is important as well to emphasize that when neuroscientists do address such questions as how neurons manage to store information, or how cell assemblies do pattern recognition, or how they manage to affect sensorimotor control, they are addressing questions concerning neurodynamics -- concerning information and how the brain processes it. In doing so, they are up to their ears in theorizing, and even more shocking, in theorizing about representations and computations. (1986: p.361)

(2) Any computational theory of cognition is constrained by neurophysiological facts. For example, she says of Marr's distinction of levels,

> [In his articulation of the parallel modelling strategy] one first had to figure out a computational schema that would solve a problem (say, for visual recognition) and only then could one usefully address the question whether and how the brain implemented that schema. Not even Marr adhered strictly to the doctrine, however, and some of the most successful parallel models are avowedly inspired and constrained by neurobiology. (1986: p.462)

(3) There is no absolute distinction between functional or computational levels and structural levels of description of the nervous system. There are many levels of organization in the nervous system; which level constitutes the functional level, and which constitutes the structural level, depends on the processes under investigation. And at many levels of organization neurophysiologists must construct computational theories defined over representations to explain how the brain performs its tasks. Hence, it is false to suppose that there is a privileged functional or computational level of

description, from which psychologists can construct their theories autonomously of neurological research.

In a sense to be made clearer below, both of (1) and (2) are true; but (3) is false. Notice first that the point Churchland makes in (3) is similar to the observation made in Section 1 of Chapter 1: In its general application, the distinction between computational and algorithmic levels of description is just the distinction between a function and its algorithmic decomposition; which description of a system constitutes the computational level of description depends on the process whose explanation we are seeking. There is in this sense no "absolute" computational level of description. However, the conclusion of Chapter 1 is that with respect to the modular processes of vision there is a distinguished computational level of description, which explains the success of visual operations, and that is autonomous with respect to the algorithmic and neurological descriptions. Thus, in vision theory at least, Churchland's assertion is incorrect.

Churchland's denial of the autonomy of the computational level misses the point that a computational-level decomposition of the relation between the image and the final representation is not strictly a description of the visual system itself. It is a description of a virtual mechanism that computes each of the relations between stimulus properties that are exploited in perception. The computational level is a description of a canonical encoding function, and of a canonical symbolic transformation function, that demonstrates how a true representation of the stimulus can be generated by mechanical operations. It is used to explain the semantic properties of visual operations via a demonstration that an algorithmic sequence yields the same output as the computational-level sequence for any given input. By contrast, the algorithmic-level decomposition is a description of the steps in the generation of the final representation by the system; there is an interpretation function from neural states of the system to the domain and range of each function in the algorithmic-level sequence. But no such claim need be made of the computational level.

Of course, Churchland is correct that the choice of computational-level theory that explains the semantic properties of actual visual processes is constrained by neurophysiological facts. A computational-level theory only describes a possible mechanism that yields veridical representations under certain conditions. Whether such a description explains the semantic properties of the actual visual system depends partially on the physical-state transition functions of the system. But at least in some cases, computational-level descriptions of the visual system can be confirmed independently of neurophysiological investigation by determining the conditions under which visual representations fail.

Recall also from Chapter 2 that a complete neurophysiological description underdetermines the choice of computational-level description. In Chapter 2 we saw that there are explanatory reasons for assigning distinct computational-level descriptions to the same algorithmic- or physical-level sequences. Thus the choice of computational-level theory for a system is partially determined by the properties of the environment in which it operates.

Thus neither Churchland's assertion that neurophysiological theory describes computations at all levels of organization, nor her point that the choice of computational theory is constrained by neurophysiological facts, demonstrates that there is no level of theory that is autonomous with respect to neuroscience. Explanations of the semantics of vision must proceed from descriptions of virtual mechanisms, formulated and selected on the basis of facts about the environment, to descriptions of algorithmic sequences that yield the same output. Hence Kitcher's claim that neurophysiology is explanatorily incomplete vis a vis the semantics of vision is correct.

As a final note to this section, we can see that Churchland's argument rests to a large extent on the view shared with functionalists that reductionism is a metaphysical issue decided by whether physicalism is true. Each of Churchland's three premises discussed above is presented as a reply to the functionalist denial of physicalism. For example, her claim that there is no distinguished functional or computational level of description of cognitive processes is offered in reply to the functionalist argument that such a level identifies a set of nonphysical properties. However, a proper denial of reductionism is not based on an ontology of cognitive processes, but rather on the explanatory considerations that determine the form of theories.

4. Optimality and idealization in computational-level theory.

In the final section of her paper, Kitcher argues that Marr's theory is an optimizing theory, i.e., it assumes that each computation is carried in an optimal way. She says,

> [Each computation] utilizes exactly the information needed for the derivation. [Marr] also assumes that, in general, the visual system is well-designed for the extraction and representation of information about the shape, spatial location, and orientation of object surfaces. (p. 21)

This conclusion is based on Marr's strategy for theory construction: One begins with a description of the information to be extracted; one then determines the information available and specifies a function that yields the information sought from the information available.

Kitcher is quite correct that computational descriptions of solutions to information-processing problems are idealized solutions expressed in canonical terms. She argues from this point that a strategy of this sort will fail to explain actual visual processes if the system does not perform computations in an optimal way. She points to the fact that some biological mechanisms are clumsy devices adapted for their particular purpose from mechanisms that once served other purposes. Here she cites Gould's (1980) description of the panda's thumb, an awkward contrivance adapted from part of the wrist bone. She claims that, while it is possible to give a description of an optimal mechanism for carrying out the task performed by the panda's thumb, such a description would be of little use in explaining how the actual thumb is successful in its function. Similarly, she argues, if the visual system is a clumsy inelegant device, a computational-level decomposition of vision will not be as useful as an investigation into the algorithms and neural processes of the system.

A very similar argument is made by Ramachandran (1985). Ramachandran suggests that, like other biological organs, the visual system may "cheat", i.e., it may use a collection of special-purpose tricks devised by trial and error. He concludes that,

> If this pessimistic view of perception is correct, then the task of vision researchers ought to be to uncover these rules rather than to

> attribute to the system a degree of sophistication it simply doesn't possess. Seeking over-arching principles may be an exercise in futility. (p. 101)

In Ramachandran's hands this argument becomes grounds for a general scepticism about the value of computational-level theories. He argues that neurobiologists ought to study the structure of the visual system and derive theories of function from these studies, rather than beginning with computational-level descriptions that ignore biological hardware. Here he compares Marr's and Gibson's approaches to perception to the four humours theory of diseases formulated by physicians who had no knowledge of internal organs.

But these claims that Marr's theory describes the visual system as employing optimal solutions to perceptual problems fail completely to grasp the role that the computational level plays in in vision theory. As we saw in Chapter 1, and in the previous section of this chapter, the computational level describes a virtual mechanism that demonstrates how information can be mechanically extracted. If, as Marr claims, early vision extracts information about the distal scene without deploying specific information about the present stimulus, there must be functional relations between higher-order properties of the stimulus resulting from natural constraints. The purpose of the computational level is to give a canonical description of these relations in a mechanically computable fashion. Thus the computational level describes facts that <u>any</u> modular process must exploit regardless of its elegance or efficiency.

The correlate of a computational-level description for the task performed by the panda's thumb is just the principles of mechanics, for it is these principles that underlie the success of the mechanism. The actual thumb of the panda, however clumsy, is successful because its construction exploits specific mechanical principles; similarly modular visual processes are successful because they exploit certain computational-level principles. Just as there is no suggestion that mechanical devices realize the idealized assumptions of mechanical principles -- frictionless surfaces, perfect elasticity, etc. --, there need be no assumption that algorithmic procedures extract all the information available as a consequence of the truth of computational-level theories. In each case, the idealizations serve to <u>generalize</u> the truth of the principles.

Ramachandran suggests that we should compare Marr's computational level to a "black box description" of the digestive system that ignores the facts about actual biological organs. But a closer parallel to the computational level in the physiology of digestion is the set of biochemical principles that underlie our explanations of the success of organs with a particular structural description in the functions they serve. So Marr's emphasis on a clear formulation of computational-level descriptions has the same force as a recommendation that physiological theories be based on a clear understanding of mechanical and biochemical principles that explain how specific organs are successful in their biological functions. Mechanical and biochemical principles are not descriptions of any specific mechanisms, yet they serve a vital role in understanding actual physiological processes. And precisely the same point applies to computational-level descriptions: Although they do not describe any specific computational mechanisms, they are essential in explaining how the algorithms employed in vision function in the extraction of information. Thus it is an error to claim that the role of the computational level is not justified unless the visual system employs optimal solutions, just as it is erroneous to claim that mechanical principles only apply to optimal mechanical devices.

Summary.

In Section 1 I argued that there is good reason to think that the denial of reductionism based on a rejection of physicalism is based on both a misunderstanding of scientific theories, and also on an incorrect view of computational explanations. The issue of reductionism is not best addressed as a metaphysical issue about the nature of mental states, but as concerning the nature of psychological method and explanation. In Section 2 we saw that Kitcher is correct that computational-level descriptions are essential to explanations of early vision, but she does not support her claim that this fact is inconsistent with reductionism. However, in Section 3 I argue that there is a basis for this claim, contrary to arguments advanced by Churchland on the grounds that neurophysiological theories include computational theories of cognitive processes. The role of the computational level is inconsistent with reduction because the computational-level decomposition is autonomous with respect to algorithmic and neurophysiological descriptions, and because neurophysiological descriptions underdetermine computational interpretations of perception.

In Section 4 I extend the points established to argue that, contrary to Kitcher and Ramachandran, Marr's theory does not depend on the assumption that visual processes are in any sense optimal. Computational-level descriptions are descriptions of relations between stimulus magnitudes, formulated in a way to show how these relations can be mechanically computed; as such they are essential to any explanation of modular visual processes. There is no assertion that actual visual processes exploit these relations in an optimal way.

CHAPTER FOUR

MODULAR PROCESSES AND THE "LANGUAGE OF THOUGHT"

0. Introduction.

In chapter 1 I compared Marr's notion of an algorithmic level of description to Pylyshyn's account of cognitive algorithms defined over expressions of a canonical "language of thought". I claimed that despite certain superficial similarities, the two accounts differ with regard to both the nature of the ideas developed and the theoretical concerns addressed. In this chapter I return to the question whether Marr's theories provide support for the assertion that cognitive processes involve a language of thought. I will argue that the considerations offered in support of the language of thought hypothesis are defences of two distinct versions of the hypothesis, one weaker than the other. My conclusion is that the strong version of the hypothesis is most likely false of early visual abilities. While there is a convincing reason for supposing that the strong version of the language of thought hypothesis is necessary to explain <u>some</u> perceptual abilities, the role played by natural constraints undermines the extension of the argument to the processes of early vision described by Marr. However, I will also argue that the weaker version is true of early vision <u>at the computational level of description</u>. In the terms developed in Chapter 1 this means that the language of thought hypothesis in its weaker form is true of the virtual machine required to explain the veridicality of early visual perception. The basis for this conclusion is a version of the second argument put forward by Fodor in the Appendix to Psychosemantics. This argument does not establish the hypothesis as an algorithmic-level assertion; nor, I think, is there sufficient independent evidence to warrant such an assertion. Since the sense in which I claim that the language of thought hypothesis is true of early vision may be viewed as Pickwickian by its opponents, in the final section I offer some reason to think that the idea of a "language of thought" does capture an important aspect of the modular systems described by Marr.

Let us delineate carefully, then, the two versions of the language of thought hypothesis.

1. The language of thought hypothesis.

The language of thought hypothesis is the assertion that explanations of cognition essentially involve a mapping from a distinguished set of neural states to expressions of a formal language. Fodor describes the hypothesis in this way:

> [Mental states] have a combinatorial semantics: the kind of semantics in which there are (relatively) complex expressions whose content is determined, in some regular way, by the content of their simple parts. (1987: p.138)

This formulation has to be broadened to incorporate the idea that, in addition to their role in bearing semantically evaluable content, symbolic expressions may encode "programs" -- sets of instructions for carrying out an algorithm. So the language of thought hypothesis asserts that there is a mapping from a distinguished set of neural states to expressions of a formal language L, and an encoding function that takes each member of L to a semantically evaluable content or to a name of an algorithmic sequence of functions.

The language of thought hypothesis is often advanced as an explication of the "representational theory of mind"; propositional attitudes thus become functional or computational relations to syntactically structured neural states. But as Fodor points out, the hypothesis does not entail the claim that the representational states it describes are relata in propositional attitudes. Thus, with regard to the case at hand, it is possible to conclude that the hypothesis is true of representational states in early vision without implying that we have <u>beliefs</u> about the states of affairs that determine the truth-values of those representational states, or alternatively that we have beliefs about the rules and procedures that are named in encodings of algorithms. Accepting both the language of thought hypothesis and the representational theory of mind commits one only to the claim that propositional attitudes are relations to a subclass of the syntactically structured representational states of the mind.

How, then, does the language of thought hypothesis apply to the processes of early vision? I have claimed that there are two versions of the hypothesis, one weaker than the other. To distinguish these two versions as they apply to early vision, it is useful to avail ourselves of a set of distinctions among computational theories formulated by Stabler (1983).

Recall from Section 1 of Chapter 1 the description of a computing mechanism. A physical system is given a computational description by specifying an interpretation function I from a set C of computational states of the system to the domain and range of a function F, the function computed by the system. Let $P: C \rightarrow C$ be a physical-state transition function that governs the behaviour of the system. Then the system computes F just in case,

$$I(P(c)) = F(I(c)).$$

According to Stabler's characterization, a theory that gives such a description of a physical system is a <u>first-level</u> computational theory. But in many cases a system computes a function F by computing an algorithm, i.e., a sequence $\langle F_1, F_2, \ldots, F_n \rangle$ of functions where the value of each function F_i, for $1 \leq i < n$, is the argument for F_{i+1}, and such that for any argument x in the domain of F,

$$F(x) = F_n(F_{n-1}(\ldots F_1(x)).$$

Stabler calls a theory that describes a physical system as computing an algorithmic sequence a *second-level* theory.

The notion of a system with a second-level theory gives us a clear formulation of the weaker version of the language of thought hypothesis. First recall the following definitions from Chapter 1: Let E be an *encoding function*, i.e., a function that maps sets of symbols to the domain and range of a function F, and let S be the set of symbol sets in the domain of E. Then a symbolic transformation function T:S --> S *realizes* a function F if, for s ε S,

$$E(T(s) = F(E(s)).$$

Now consider a mechanism M in the cognitive system that computes a function F. The *weaker version of the langauge of thought hypothesis* is true of M just in case (1) M computes F by computing a symbolic transformation function T that realizes F, and (2) M has a second-level theory according to which it computes T by computing a nontrivial algorithmic sequence $<T_1, T_2, \ldots , T_n>$ for T, where each T_i realizes a corresponding function F_i in an algorithmic decomposition $<F_1, F_2, \ldots , F_n>$ of F.

Let us see how the weaker version of the language of thought hypothesis can be applied to the explanatory framework of Marr's empirical theories outlined in Section 2 of Chapter 1. Recall from that discussion that, at the computational-level, the processes of early vision are described as computing a sequence of functions that yield successive representations of the stimulus. We also saw, however, that his assertion does not entail the claim that at the algorithmic level the system computes each function in the sequence. Let us say that the weaker language of thought hypothesis is true of the early visual system if the system has a second-level theory according to which, for each function F_i in the computational-level description, the system computes a symbolic transformation function T_i that realizes F_i.

The stronger version of the language of thought hypothesis asserts that a system is, in Stabler's terms, a "program-using system". The notion of a program-using system is defined in the following manner. In many cases, a system computes an algorithmic sequence because its behaviour is governed by an internal representation of a program, i.e., a set of instructions for carrying out the sequence of steps in the algorithm. A program can be thought of as a set of symbolic formulae that encode a sequence $\langle IN_1, IN_2, \ldots IN_n \rangle$ of instructions, where IN_i is an instruction or command to compute a function F_i in an algorithmic sequence. Then a system is program-using just in case there is a "program realization" function that maps the encoded instructions to physical states of the system, together with a set of "control states" associated with the encoded instructions such that the system computes the function F_i specified by IN_i whenever it is in the control state associated with that instruction. Stabler calls a theory that describes a system in this way a third-level theory.

Applied to the visual system, the stronger language of thought hypothesis asserts that (1) the weaker version of the language of thought hypothesis is true of the visual system, (2) physical states of the visual system encode a program that specifies the symbolic transformation functions that generate the representations in the sequence, and that (3) the system computes the sequence of symbolic transformation functions because its operation is governed by the encoded program.

The strong version of the language of thought hypothesis is defended, with respect to cognition generally, by Pylyshyn (1983). According to Pylyshyn, the appropriate mathematical model for cognition is an infinite automaton, i.e., a computing device with no finite upper bound on the number of distinct states in its memory.[1] And this claim is equivalent to the assertion that the cognitive system is a program-using system. Let us see why this is so.

---

[1]. The class of infinite machines thus includes among its members both push-down automata and Turing machines.

The essential feature of a finite automaton is that its potential memory is exhausted by the set of states in its machine-state diagram; i.e., given a complete specification of the machine-state diagram, the final state of a finite automaton is determined by its present machine state and its unread input. By contrast, the final state of an infinite automaton is determined by its present machine state and its unread input, together with what is written on its tape or register. It is the fact that it has access to a tape or register with a syntactic structure that gives an infinite automaton an unbounded memory; for, given this structure, the set of distinct possible states of the tape or register is recursively generable. Hence the essential feature of infinite automata is that their operation is governed by internal states with a syntactic structure, and it is just this feature that distinguishes program-using systems. So the stronger version of the language of thought hypothesis is true of the cognitive system only if its most appropriate mathematical model is an infinite automaton.

By contrast, the weaker version of the language of thought hypothesis allows that the most appropriate model of the system is a finite automaton. This version is defended, with respect to modular systems in language and vision, by Fodor. In his response to Stabler's scepticism over the hypothesis that transformational grammars are explicitly represented in neural structure, Fodor gives the following argument.

> RTM says that the contents of a sequence of attitudes that constitute a mental process must be expressed by explicit tokenings of mental representations. But the rules that determine the course of the transformation of these representations ... need not themselves ever be explicit. They can be emergents out of explicitly represented procedures of implementation, or out of hardware structures, or both. Roughly: According to RTM, programs -- corresponding to the 'laws of thought' -- may be explicitly represented; but 'data structures' -- corresponding to the contents of thoughts -- have to be. (1987: p. 25)

He completes the point in this way:

> Restricting one's attention to the status of rules and programs can make it seem as that the computer metaphor is neutral with respect to RTM. But when one thinks about the constitution of mental processes, the connection between the idea that they are computational and the idea that there is a language of thought becomes immediately apparent.

This claim is explicitly asserted of vision theory in the Appendix to Psychosemantics. (1987: p.144-145).

Although these two versions of the language of thought hypothesis are distinct assertions, arguments for one version have been placed alongside arguments for the other. Part of the reason for this is that, as in the quotation above, the arguments for both are usually applied to the question whether perceptual mechanisms are "representation-using", and this term is often used in a way that collapses several important distinctions. There is a perfectly natural sense in which the phrase "representation-using" can refer simply to a system that has a principled computational description, i.e., a system for which a computational theory has genuine explanatory value. But as we will see in the next section, according to Fodor a system is representation-using just in case it satisfies the weaker version of the language of thought hypothesis. And the term is also occasionally employed in conjunction with the term "representation-governed", which usually means that the system is program-using.

There is one argument in particular, which defends the language of thought hypothesis with respect to perceptual abilities, that has generated just such a confusion of issues. Fodor (1986) claims that the postulation of mental representations is necessary to account for our ability to respond selectively to "nonprojectible properties", i.e., those properties that do not enter into natural laws. We will see in a later section that this argument does provide an explanatory basis for computational explanations, and moreover that it gives prima facie reason to suppose that some perceptual systems are program-using. We will also see that when properly understood it does not apply to the processes of early vision described by Marr. However, the argument can be read in at least two ways. And Fodor's descriptions of the arguments suggest an unsuccessful line of reasoning to the weak version of the hypothesis. Since I argue below that the language of thought hypothesis in its weaker form is true of early vision at the computational level, it is important to distinguish the reasons for this conclusion from the unsuccessful version of the argument from nonprojectible properties. This is especially so since remarks of Fodor's suggest that the argument from nonprojectible properties is of a piece with what I see as the correct reason for asserting the weaker language of thought hypothesis of early vision. Accordingly, let us first isolate the bad version of the argument from nonprojectible properties; we can then see more precisely why the better version of the argument does not extend to early vision.

2. One argument from nonprojectible properties.

Fodor proposes that a system is representation-using just in case it responds to "nonprojectible" properties of the stimulus, properties that do not occur in any natural laws. Since there are properties of this sort that we do perceive, we can infer that some perceptual processes are representation-using in this sense. The question, of course, is whether this notion captures what is intended by some natural use of the term "representation-using". For this claim Fodor offers the following argument: Suppose that a system S responds selectively to a nonprojectible property O of an object A. There must be some nomological relation between S and and a property of A that makes this possible, but since O is nonprojectible _it_ cannot enter into such a relation. The detection of the relevant projectible properties must eventuate in S coming to represent A as being O, and this process must be a case of perceptual inference -- typically an inference that these properties co-occur with O.

Fodor's presentation of the argument from nonprojectible properties is based on his general strategy for explicating computational explanations. According to this strategy, a system is representation-using just in case its state-transition map cannot be expressed in terms of a nomological relation between physical variables. The argument from nonprojectible properties then takes the following form: Suppose a system S responds selectively to a nonprojectible property O. Since there are no laws in which O appears, perception of O cannot be explained in terms of nomological relations between O and states of S. But a system is nonrepresentation-using only if explanation of its behaviour is capturable in terms of nomological relations between states of the system. Since this is not the case with respect to the perception of O, S is representation-using.

Fodor intends this argument to provide support for the weak version of the language of thought hypothesis. Thus, according to Fodor, a system is representation-using just in case it computes a sequence of symbolic transformation functions. Moreover, according to

Fodor the weaker version of the language of thought hypothesis is constitutive of the principled application of computational descriptions; i.e. a computational description has genuine explanatory value only if the system computes a sequence of symbolic transformation functions. In contrast to explanations of perception by representation-using systems, Fodor describes what he calls a "primal scene of the first type", the defining characteristic of which he gives as the occurrence of "a lawful connection between a property of the 'stimulus' (viz. S's coming to be O) and a property of the ensuing behavioural response (viz. A's behaviour coming to be C)." What Fodor seems to have in mind is what I have called "direct detection", where the explanation of the covariance between a stimulus magnitude and states of the system involves descriptions only of properties of the magnitude itself and of the physical properties of the system.[2] If so, then on Fodor's view, any perceptual system whose explanation essentially involves a computational theory is a representation-using system in his sense.

But the strategy upon which the argument is based is not a good one. To see this, and to see how Fodor identifies the weaker version of the language of thought hypothesis with the principled application of computational descriptions, let us look at the strategy in some detail.

---

[2]. Chapter 1, Section 3.

The foundation of the view that computational descriptions provide genuine explanations of mental processes is that, in a sense not yet fully understood, they enable us to express true generalizations that cannot be expressed in the vocabulary of physics or neuroscience. One way of trying to explicate this idea is to assert the existence of a general class of mechanisms that exhibit behavioural regularities that cannot be expressed in noncomputational terms. Any device that operates according to physical laws has a computational description, but in many cases such descriptions capture no behavioural regularities that are not expressible in a physical characterization of the device. On this view, understanding the explanatory power of computational descriptions is to be sought by establishing the nature of the distinction between these two classes of mechanisms. Moreover, it is widely agreed that the behavioural regularities that computational descriptions capture are those that are stated over the syntactic and semantic properties of representational states. Hence the distinction that constitutes the basis of computational descriptions can be grasped by formulating a distinction between representation-using and nonrepresentation-using systems.

Several proposals within this program identify the distinction between representation-using and nonrepresentation-using systems with the distinction in computer engineering between digital and analogue computers. The motivation for this is clear: As commonly understood in computer engineering, digital computers operate on discrete symbols. However, the use of a discrete form of representation is evidently not a sufficient condition for digital computation, for we can easily construct intuitively analogue devices that work in discrete steps. According to Fodor and Block (1973), the distinction between digital and analogue machines lies in the nature of the relationship between the representing magnitudes rather than in the nature of the representing magnitudes themselves. On their view, the relevant feature of analogue computation is the nomological character of the relation between the physical variables used in the computation. Thus they call a mechanism analogue if the state-transition map is a physical law, and digital otherwise. The point is echoed by Pylyshyn:

> The significance of the nonprojectibility of the class of physical properties corresponding to distinct computational states of a digital computer is that the operation of the system, as a computer rather than a physical system, cannot be explained the way natural events typically are explained in science, by citing the value of the property in question and showing that the state transition is subsumed under some general, natural law. (1984: p.201)

Systems that operate on symbolic representations, it is argued, depend on the preservation of appropriate relations between symbols, and these relations are expressed in terms of the syntactic and semantic properties of the system rather than the physical laws that underlie the symbolic transformations. Thus it is clear how this strategy for explicating computational theories is seen to support the view that the weak version of the language of thought hypothesis captures the explanatory significance of computational descriptions. For any system that uses symbolic-transformation functions depends on those functions preserving the appropriate relation between symbol tokens; hence, so the argument goes, a system has a principled computational theory just in case it computes a sequence of symbolic-transformation functions.

However, as Demopoulos (1987) points out, one cannot draw the distinction between representation-using and nonrepresentation-using systems on grounds of the nomological character of the state-transition map. The behaviour -- computational or otherwise -- of any machine is never simply a consequence of a physical law, but of the law together with a description of the constitutive structure of the mechanism. In the computational case this requires the specification in physical terms of the computational states and the transitions between them. But given its constitutive structure every device obeys some family of physical laws. Hence Fodor and

Block's proposal is either vacuous or reduces to a pragmatic distinction between simple and complex laws.[3]

The strategy of drawing the distinction between representation-using and nonrepresentation-using systems along nomological-nonnomological lines is tied to Fodor's view, described in Chapter 2, that computational descriptions appeal to causal properties of states partitioned by their representational content. The generalizations that are expressed by computational theories assert that diverse physical systems share the same computational description. But on Fodor's functionalist explication of computationalism, classes of computational states that share a syntactic or semantic interpretation are functional states. And, since on his view functional states are not identical to physical states, it is argued that their causal properties are not subsumable under a physical law. Thus, according to Fodor, the explanatory value of computational theories lies in the fact that they capture relations between functional states, which cannot be reduced to nomological relations between physical states. But as we have seen in earlier chapters, the functionalist explication of computationalism, whereby

---

3. According to Demopoulos, the analogue-digital distinction can be an interesting one when it is applied to classes of mechanisms: A class of mechanisms that compute a given function is analogue if the set of their state-transition diagrams has a unitary physical description, and digital otherwise.

computational theories describe causal relations between functional states, is a mistaken view of the form of computational theories. The generalizations that are captured by computational theories are not causal generalizations; rather they are assertions that a class of physical systems is equivalent under a computational description.

However, this is not the only interpretation of the argument from nonprojectible properties; let us look at the other, better, version.

3. A better argument from projectible properties, and its application to early vision.

Matthews (1986) offers the following criticism of the argument from nonprojectible properties. Matthews argues that there are many examples of intuitively nonrepresentation-using devices that respond selectively to nonprojectible properties. We need only consider a device that detects a range of projectible properties P_1, P_2, ... P_n, where these co-occur with a nonprojectible property P^*. Matthews argues that, whatever interpretation is placed on the term "infer", we can maintain that such a device infers the presence of P^* from the presence of P_1, P_2, ... P_n. As an example of such a device, Matthews offers a description of a simple mail sorter: P^* varies over zip codes, and the sorter responds selectively to projectible properties of printed characters.

Notice that Matthews' criticism is effective against the version of the argument from nonprojectible properties described in section 2. That version claimed that the presence of P^* is inferred because there are no laws in which it occurs. But the relevant laws that occur in any explanation of perception are those involving the projectible properties of each instance of P^*; hence in each case, given the constitutive structure of the device together with a physical description of the possible instances of P^* to which it responds, there are nomologicals that express the relation between the behaviour of the device and the particular occurrence of P^*. These nomologicals may well be complex and unnatural; but to base the response to Matthews' criticism on these grounds would render the distinction between representation-using and nonrepresentation-using mechanisms arbitrary.

Demopoulos offers the following reply to Matthews: The class of properties to which the mail sorter described by Matthews responds may have a "unitary physical description"; i.e., there may be a single well-motivated, or natural, description of the class in the language of the physical sciences. If the class of properties to which a mechanism responds has a unitary physical description, then describing its behaviour as responding to a nonprojectible property is merely a convenience. But zip codes have many possible physical realizations, and the device may be constructed in such a way that there is no unitary physical description of the class to which it is responsive. In this case the principle upon which the behaviour of the device is based comes from outside physical science. We can say that the class of properties to which a device of this sort responds is "open-ended" with respect to physical sciences. And if the class to which it responds is open-ended in this way, the description of the mechanism as responding to a nonprojectible property is indispensable since it provides the only generalization that covers all members of the class to which the device is responsive.

Demopoulos' response to Matthews is intended to provide a basis for attaching explanatory significance to computational descriptions. But as Demopoulos points out, response to nonprojectible properties under an open-ended class of physical realizations is not a <u>necessary</u> condition for principled application of computational descriptions. Consider a device which, like Matthews' mail sorter, responds selectively to a class of stimulus properties that has a unitary physical description. It may be that there is no explanatory gain in a computational description of the mechanism. The properties may be <u>directly detected</u> in the sense that the explanation of the mechanism simply involves a description of the stimulus properties together with a description of the physical processes of the mechanism. But this need not be the case. It may be, for example, that the device computes a relation between the properties in question and some other set of stimulus properties. This is precisely the explanatory basis of Marr's computational theories. There is no reason to suppose that the stimulus properties represented by states of the system under Marr's descriptions co-occur with open-ended sets of projectible properties; indeed they are themselves plausibly projectible properties. The explanatory value of computational descriptions in these theories lies in an appeal to the fact that physical-state transitions of the mechanism belong to a class of computationally equivalent systems that compute relations among stimulus magnitudes.

Neither does the argument from nonprojectible properties by itself support either version of the language of thought hypothesis. According to Fodor, response to a nonprojectible property must involve symbolic transformation functions. As Fodor sees it, the basis of appeal to computational descriptions in Marr's theories is the same as that in cases of response to nonprojectible properties. In each case, the explanatory value of the theory is held to rest on causal generalizations stated over symbolic representations. But we have seen that this view rests on his mistaken account of computational theories. A mechanism that responds to an open-ended class of physical properties has a principled computational description, but there is nothing in this idea that entails the claim that the computation involves symbolic transformation functions.

However, if the range of physical realizations of a nonprojectible property to which a device is responsive is sufficiently large, there is reason to suppose that the system is program-using. Let us look at this carefully.

Consider a device that responds to a nonprojectible property under an open-ended class of physical realizations. The responsiveness of the device in the presence of each individual realization may not require a computational description; the explanatory value of the computational description may lie simply in its ability to collect diverse physical behaviours under a common description in terms of the nonprojectible property. In such a case there is no basis for either version of the language of thought hypothesis. On the other hand, the responsiveness of the device may involve computation of relations between properties of the stimulus. While the account of the device in the presence of each realization requires a computational description, its behaviour may not require greater resources than a finite automaton. In principle, the computational resources of a finite automaton will not be exceeded as long as the number of distinct internal states of the device required to explain its input-output behaviour for all possible stimuli is finite.[4]

However, if the number of distinct physical realizations of a nonprojectible property is sufficiently large, it is likely that the system is program-using. For when the range of factors that affect the response of the device to a given physical stimulus is sufficiently large there is reason to suppose that the system exploits combinations of stored algorithmic sequences. According to Pylyshyn, the number of distinct factors that may influence response to the perceptual categories of propositional-attitude psychology is infinite. If this is the case, then appeal to the computational resources of a finite automaton is insufficient to model the computational behaviour of the cognitive system. For the number of distinct internal states of the system required to explain its response to physical realizations of nonprojectible properties will be infinite. The intuitions here are not completely clear, but the arguments are sufficiently strong to lend prima facie support to the claim that human cognition at some level of organization is program-using.

---

[4]. By this I mean only that its input-output behaviour can be mimicked by a finite automaton, not that its internal architecture conforms to the standard formal description of finite automata.

This can be the case, it seems, even if the class to which the device is responsive is infinite. Here the intuition is based on the ability of finite automata to recognize infinite languages; this is possible when strings in the language contain repeatable substrings so that the algorithm contains a loop or cycle.

Let us see how this applies to early visual perception. The role of natural constraints in computational theories of early vision is to restrict the information accessible at each stage of the process while ensuring the normal veridicality of the output. But the explanatory advantage in appeal to program-using systems is to explain the influence of a wide range of factors on the response of the mechanism. Hence it is reasonable to expect that systems whose explanation involves appeal to natural constraints will be those that generate information about the stimulus without recourse to symbolic memory. This suggests a division of perceptual processes into two classes along the lines of Pylyshyn's distinction between operations of the functional architecture and program-using systems. But, as we will see in the next section, the processes postulated to explain the role of natural constraints and the processes that constitute Pylyshyn's notion of the functional architecture do not necessarily coincide.

First let us review what we have established so far. We have seen that the question whether perception involves a language of thought actually embraces two issues: (1) whether visual processes compute symbolic transformation functions; (2) whether the visual system is program-using. And we looked at two versions of the argument from nonprojectible properties: The first claims only that perception involves symbolic transformations, but we have seen that it is based on an erroneous program for explicating computational theories. According to this program, the explanatory value of computational theories lies in the fact that the state-transitions of computational systems are not expressible as nomological relations; the argument from nonprojectible properties then asserts that perception of nonprojectible properties is not explicable in terms of physical laws. The second version of the argument offers prima facie reason to suppose that some perceptual processes are program-using. We established that the processes described in Marr's theories are naturally excluded from this set because the role of natural constraints is precisely to eliminate the kind of plasticity of behaviour that program-using systems make possible.

In the next section I will consider the fact that the line of reasoning that led to program-using systems in perception may exclude the processes described in Marr's theories from even the <u>weaker</u> form of the language of thought hypothesis. In Section 5 I will look at an argument of Fodor's to the effect that the weaker form of the hypothesis is true of early visual processes.

4. Functional architecture and symbolic transformations.

In the discussion of Pylyshyn's notion of functional architecture in Chapter 1, it was noted that the claim that the cognitive system has an internal syntax imposes a natural distinction between two kinds of symbolic operations. There are the processes that are governed by internal symbolic states, which on Pylyshyn's view are those that are stated over the categories of propositional attitude psychology. There must also be a set of operations that effect the transformations among symbols, and that determine which transformations to execute at any one time. The latter set comprises what Pylyshyn refers to as the functional architecture of the system. The operations of the functional architecture compute symbolic transformation functions, but on pain of circularity they must not themselves involve internal symbolic states. So the functional architecture includes computational operations whose initial and final states are symbol sets but whose most appropriate model is a finite automaton. We must widen this description, however, to include mechanisms that translate physical stimulus information into symbols, and mechanisms that take symbolic output to physical behaviour.

According to Pylyshyn, the chief characteristic of operations of the functional architecture is that they are "cognitively impenetrable"; i.e. their input-output behaviour is unaffected by changes in the subject's background beliefs. And it seems clear that cognitive impenetrability is a necessary condition for membership in the functional architecture, on the grounds that processes can be influenced by background beliefs only if they are governed by internal representations. But is there a finer-grained decomposition of the system into primitive symbolic transformation functions than the one generated by the stipulation that a function is primitive if it is not governed by internal symbolic states? Some of Pylyshyn's comments suggest a negative answer. And so far we have provided no criteria for concluding that a system involves symbolic transformation functions except that we have reason to believe that its behaviour is governed by internal symbolic states. That is, the only possible evidence we have discussed for asserting the weaker version of the language of thought hypothesis of a system is evidence that it satisfies the stronger version of the hypothesis.

From what we have seen, it does not seem likely that program-using systems are required for any of the computations in early vision. So if the set of primitive symbolic transformation functions of early vision is the set of functions that are not governed by internal symbolic states, it is likely that the entire sequence from the grey-level image to the final representation is a single primitive step. On the other hand, Pylyshyn suggests in some places that it is the individual functions in the computational-level sequence from the grey-level image to the final representation, rather than the sequence itself, that comprise the functional architecture of early vision. (See for example 1984: pp. 214f.) But so far we have no basis for this assertion.

5. A language of thought at the computational level.

There is an argument by Fodor (1987: pp. 143-147) that offers one good reason to assert that the weak version of the language of thought hypothesis is true of early vision. Fodor's point is that if we assign the representational content of computational states to sets of symbols, then manipulation of the symbols provides a mechanical way to generate one representation from another. Here is how Fodor puts it.

> Mental symbols constitute domains over which mental processes are defined. If you think of a mental process -- extensionally, as it were -- as a sequence of mental states each specified with reference to its intentional content, then mental representations provide a mechanism for the construction of these sequences; they allow you to get, in a mechanical way, from one such state to the next by performing operations on the representations. (p. 145, emphasis in the original.)

There is an important sense in which this argument is sound. In particular, the argument can be applied to early vision as the assertion that we can only arrive at descriptions of physical processes that generate true representations from other true representations by basing the description on a computational-level description of a sequence of functions defined over symbol sets. By devising formal schemes we give precise descriptions of the information-content that representational states are held to carry. Then, as long as the symbolic transformation functions appeal only to syntactic properties of the symbol sets, we know that the mapping from one representation to another preserves (by physically realizable operations) all the information claimed for it. This is in fact just a restatement of methodological solipsism.

From what we have seen in earlier chapters, it is clear that Fodor's argument succeeds in establishing that at the computational level, visual processes can be modelled only on the basis of a decomposition of the relation between the grey level image and the final representation into a sequence of functions defined over symbolic representations. But this does not preclude the possibility that at the algorithmic level the function from image to final representation is primitive in the sense that there is no algorithmic decomposition of the relation into operations on symbol sets.

However, in favour of such an algorithmic-level decomposition there are the kinds of prima facie considerations offered by Ullman (1980) based on the complexity of the relation. Ullman points out that the most immediate difference between Gibson and Marr on the correct form for theories of perceptual abilities is Marr's incorporation of an algorithmic level of description. According to Ullman, visual perception is direct in the way Gibson intends only if the pick-up of information is psychologically primitive in the sense that there is no decomposition of the operation in terms of how the information is represented by internal states of the system. If Gibson is correct, for each stimulus there must be a mechanism tuned to it that produces a specific percept, where the algorithmic decomposition of this mechanism is possible only in neurophysiological terms.

To demonstrate the implausibility of Gibson's theory, Ullman compares Gibson's and Marr's theories to alternative explanations of a computing device that performs integer addition. If the number of calculations such a device is designed to perform is small, the calculator may simply use a direct pairing of inputs and outputs. In this case there is no decomposition of the operation in algorithmic terms; describing how the function is computed is simply a matter of providing an interpretation of the circuit diagram. But if the device is designed to perform additions over a large number of input pairs, the operation may implement the standard algorithm which breaks the operation into a sequence of sums of powers of a base number. Understanding how such an operation is implemented in the physical structure of the calculator requires knowledge of how the integers are represented as well as knowledge of the algorithm defined over the representations.

Ullman argues that the same considerations apply to understanding visual perception. In vision theory we want to explain how neurological processes generate sensory representations, or percepts, of the external world from changing states of the retinal image. According to Ullman, given the complexity of the function from stimulus patterns to the final representation it is unlikely that the visual system uses a direct pairing of initial and final states in the manner of the simple calculator. It is much more reasonable to suppose that percepts are computed by implementing a set of formation rules, defined over primitive symbols, that construct complex representations. In this case we can only understand how the visual system operates by determining the representations and algorithms involved.

But given the increased interest in the computational power of parallel distributed architectures, it is not well established that there is an algorithmic level decomposition of the kinds of sequences described by Marr. It is these concerns that cause doubts about the language of thought hypothesis in its weaker form. So while Fodor is perfectly correct about the need to characterize perceptual processes as symbolic transformation functions in order to establish the possibility of mechanical realization of mappings with the right semantic properties, we must not confuse this point with a claim about the actual realization implemented in the system.

6. Why a language of thought?

Before closing I would like to address the question whether, given the considerations above, there remains any natural sense in which the processes described by Marr employ a "language of thought". As its name suggests, the language of the thought hypothesis makes two distinct assertions. One is that there is an internal language, and the other is that it is a language of thought. The problem with asserting the first claim with respect to early visual processes is that the formal schemes framed in the theories are not very much like languages; in particular, they do not have the usual recursive properties of natural languages. Recall that a formal scheme involves only a mapping from neural states to sets of symbols and the specification of a symbolic transformation function; the class of symbols sets that such a scheme employs need not need be recursively generable, nor need individual strings of symbols contain embedded expressions of arbitrary length, as one finds in natural languages. But as long as we are clear about the differences between formal schemes and languages in the usual sense, this is perhaps not so important. Does the reference to thought have a natural motivation? Despite the fact that the language of thought hypothesis does not entail that we have beliefs about the states of affairs represented, it may still be objected that reference to a language of thought in early vision implies that mentalistic terms, like "inference" and "reasoning", apply properly to modular processes where they are clearly inappropriate.

The aversion to the use of mentalistic terminology is that appeal to these terms is sometimes taken to rest on the fact that states of the system are representational. The argument moves from the claim that visual states are representational to the assertion that the processes that generate them can be seen as realizing relations between intentional states. But this argument leads us, by parity of reasoning, to the conclusion that mentalistic terms apply to the explanation of thermostats, paramecia and servomotors. This kind of concern is apparent in Matthew's criticisms of Fodor's program.

Against this concern, it is usually argued that mentalistic descriptions apply more naturally to program-using systems. For in descriptions of these systems mentalistic terms apply to features of the behaviour of the system that do real explanatory work in the theory: There is evidence that a system is governed by symbolic internal states when the appeal to inference captures behavioural generalizations unavailable to other theoretical descriptions. The problem with applying mentalistic terms to thermostats and paramecia is that there is no aspect of their behaviour, to which the terms might apply, that plays any role in explaining the covariance between states of the system and the represented magnitudes. But since we cannot rule out the possibility that all of early vision is carried out in a single algorithmic-level step, on this line of reasoning the use of mentalistic descriptions for these processes seems misleading.

However, the use of mentalistic terminology is not without any basis in early vision. My argument here is that there is a natural sense in which such systems are susceptible to error. Thermostats and paramecia are systems that "directly detect" the magnitudes with which they covary; i.e., the covariance is explained by appeal only to properties of the magnitudes represented and to the physical properties of the system. For these systems, failure of covariance occurs only when the system fails to operate normally. The failure is explained by the fact that the nomological relations between states of the stimulus and states of the mechanism are different than they are when it truly represents the stimulus magnitude. In such a case, there is no natural sense in which we can say that the system infers, say, the temperature of the room from the state of a bimetal strip. For there is no possibility that the system would behave in the same way and <u>fail</u> to represent the magnitude in question. Thus we can describe the system in representational terms only in circumstances in which it operates veridically; there are no cases in which the system can be said to falsely represent a particular magnitude.

It is more reasonable to describe the operation of modular computational systems that exploit natural constraints as inferential because they compute functions from representational content to representational content where failure to represent veridically is explicable as a consequence of the normal operation of the system. Circumstances in which the system fails to covary with the relevant stimulus magnitudes are specified over the relations between states of the system that define its computational behaviour; the system need not malfunction in order to fail to covary with a stimulus magnitude. Hence we can assign representational content to states of the device independently of whether the representation is true. So computational-level descriptions of the sort described by Marr are perhaps the simplest cases in which we can meaningfully apply the notion of error.

7. Summary.

We can now summarize the conclusions reached in this chapter with respect to a language of thought in early vision. Of the two versions of the language of thought hypothesis, there is no basis for claiming that the stronger version is true of early visual processes; that is, it is probably false that the early visual system is a <u>program-using system</u>. In fact, modular processes of the sort described by Marr are those that are designed to avoid the kind of plasticity of behaviour that makes program-using systems useful. However, there is a theoretical justification for claiming that the weaker version -- the claim that visual processes compute symbol transformation functions -- is true of early vision <u>at the computational level</u>. While there is an argument, based on the complexity of the relation between the image and the final representation, to the conclusion that the weaker version is true at the algorithmic level also, the current evidence is insufficient to draw firm conclusions.

Since there is not enough evidence to assert even the weaker version of the hypothesis to early vision, it appears unreasonable to claim that there is a language of thought in any interesting sense in early vision. This conclusion is justified, I think -- especially with regard to the claim that the symbolic transformations of vision constitute a <u>language</u>. Yet the fact that the notion of error has a genuine explanatory role lends some support to the use of mentalistic terminology in descriptions of the modular processes of vision.

APPENDIX ONE

GIBSON'S THEORY OF DIRECT PERCEPTION

0. Introduction.

In Chapter 1 I noted the close relation between Gibson's and Marr's use of global properties of the natural environment in explaining the veridicality of perception. In Section 1 of this appendix I discuss Gibson's arguments against computationalism, which he bases on this feature of perception theory. I conclude that Gibson's anti-computationalist claims are based on a misunderstanding of the significance of higher-order variables in perception. In Section 2 I argue that Gibson's claim that perception theory should not be based on experimentation involving illusion is similarly flawed.

1. Mentalism and higher-order variables.

According to Gibson, the important aspect of the idea of higher-order variables in the ambient light is that it demonstrates how perception can be explained in terms of relations between physical variables.

> The inhomogeneities of the retinal image can be analyzed by the methods of number theory and modern geometry into a set of variables analogous to the variables of physical energy. This says, in effect, that the order or pattern of the retinal image can be considered a stimulus. (1955: p.9)

The basis of this claim is that values of higher-order variables need not be computed from first-order values but can be directly detected by an appropriately constructed device. By analogy, Gibson points to the manner in which radio receivers are tuned to a particular frequency; there is no sense in which the receiver internally represents the frequencies of incoming signals and selects the correct one. Gibson suggests that in a similar way the visual system is "tuned" to pick up the values of higher-order variables in the ambient light. There is no need for the system to calculate higher-order values from the values of light stimulation at individual points on the retina as, on Gibson's account, classical theories have traditionally assumed. According to Gibson, once the information content of patterns in light stimulation has been determined, all that remains for a complete theory is the isolation of neurophysiological structures that respond to appropriate patterns.

Part of Gibson's argument against internal representations is based on his identification of computations on interpreted states with unconscious inference from momentary sensory states. Thus, on his view, the appeal to representations in perception theory is a species of "mentalism", i.e., the claim that cognitive processes should be explained in terms of the categories of rational thought. Gibson argues that historically the failure to recognize the role of global regularities in the physical environment in fixing visual information led both empiricists and nativists to search for processes that supplement the meagre information supplied by two-dimensional point-stimulation at the retina. As he sees it, the only premise supporting mentalism is that three-dimensional properties of objects cannot be determined from momentary point-stimulation, and so must presumably be inferred with the aid of background knowledge, either innate or learned. Hence, recognition of information carried by higher-order variables eliminates the support for mentalism. A theory of visual perception, as he sees it, has only two components: (1) a description of the information available in the spatial structure of ambient light, and (2) a neurophysiological theory that explains how this information is picked up by the visual system. There is no need on this view for a description of how the information is represented and processed by the organism. On his view, it follows by the same reasoning that descriptions of perceptual processes as <u>computational operations</u> on representations are fundamentally misguided. He says for example,

> Not even the current theory that the inputs of the sensory channel are subject to "cognitive processing" will do. The inputs are described in terms of information theory, but the processes are described in terms of old-fashioned mental acts: recognition, interpretation, inference, concepts, ideas, and storage and retrieval of ideas. These are still the operations of the mind upon the deliverances of the senses, and there are too many complexities entailed in this theory. It will not do, and the approach should be abandoned. (1986: p.238)

However, Gibson's argument against internal representations in computational theories rests on a misunderstanding of the true significance of higher-order variables. The role of higher-order variables does not eliminate appeal to computational relations among representations. Rather relations between higher-order variables in the stimulus permit the use of modular computational systems. As we saw in Chapter 4, Gibson's claim that early vision does not have a decomposition into representational processes may be correct at the algorithmic level. But this is not for the reasons that he cites in his criticisms of the computationalist program. On Gibson's view, internal representations are postulated to explain how impoverished stimuli are supplemented by background information from memory to produce veridical perception. Once this requirement is dropped, he argues, the appeal to representations is without basis. But questions concerning the veridicality of early perceptual processes are computational-level issues involving the information-processing problems involved in explaining perceptual behaviour. Considerations involving the character of internal representations and processing are algorithmic-level questions which concern the realization of computational-level functions, and not the supplementation of impoverished stimuli as Gibson supposes.

2. Illusions and natural constraints.

The discussion so far reveals the error in Gibson's insistence that theory construction should not be based on experimentation involving perceptual illusions. According Gibson, the two requirements: (1) that a theory specify the conditions under which perception is nonveridical; and (2) that it provide explanations of these perceptual failures, are both misguided. Gibson argues that theory construction should be an attempt to uncover higher-order variables in the image that are correlated with properties of the physical world, and thus serve as channels for information about the distal scene. On his account, experimentation involving illusions is based on the assumption that illusions reveal the nature of the internal representations involved in unconscious inference. These representations, he claims, are hypothesized on false grounds. Hence perception theory ought to concern itself with the conditions under which perception is successful. Experimentation involving illusions, so the argument goes, only leads to failure to identify these conditions.

While Gibson is correct that the first task of theory construction is to identify the stimulus relations utilized in obtaining information about the world, knowledge that a given variable <u>can</u> provide information about the distal scene must be coupled with a confirmation that the variable <u>is</u> used in this way by the system. Thus even at the outset, experimental procedure must involve removing potential sources of information to isolate those that are actual sources from those that do not affect the performance of the system. Discovering the conditions under which illusions occur is the only way of ascertaining the particular computational-level theories implemented in the operation of the system.

Gibson's standard reply to arguments of this form is that illusions are situations in which the full higher-order variables of normal perception are not presented to the system. Hence, he argues, illusions do not properly identify those variables. For example, stereograms give illusory perceptions of depth cnly because some part of the complete stimulus array is absent; hence veridical binocular perceptions of depth are the effect of a more complex variable in the stimulus than that available in a stereogram.

A problem with this reply was first formulated in Chomsky's (1959) critique of Skinner; that it applies to Gibson's theory has been observed by Fodor and Pylyshyn (1981).[1] The problem is that without constraints on descriptions of higher-order variables, explanations become trivialized in the sense that whatever stimulus

---

[1]. As Fodor and Pylyshyn note, Gibson does not commit the same error that Chomsky attributes to Skinner. According to Chomsky, Skinner collapses the distinction between controlling stimulus and the property perceived. By contrast, Gibson does distinguish distal properties from their corresponding stimuli, but his argument is nonetheless similar in form.

array causes veridical perception of a feature of the stimulus serves to define the higher-order variable for that feature. According to Fodor and Pylyshyn, we need an independent criterion for what counts as an appropriately bounded segment of the stimulus array. (1981:p. 171) But while interesting, this line of criticism misses the critical error in Gibson's theory. For, as we saw in Chapter 1 and in the last section, Gibson's emphasis on the identification of higher-order variables is a red-herring. The task of perception theory is not just to identify the variables exploited in veridical perception, but also to supply a complete explanation of how the response to such a variable produces veridical perception. According to both Gibson and Marr, perception of most features of the distal stimulus depends on a relation between it and some other feature of the stimulus – one that is a consequence of a specific set of natural constraints. Experimentation involving illusions is necessary to identify those natural constraints, and thereby to identify the relations between physical magnitudes that support perception of the distal stimulus.

APPENDIX TWO

STEREOPSIS AND THE MATCHING PROBLEM

In Chapter 1 I argued that the computational level of description of early vision is autonomous of the algorithmic decomposition of the system in the following sense. The functions that appear in a complete computational-level description of the relation between the image and the final representation are determined by the relations between stimulus magnitudes upon which the success of visual operations depends. In this way the set of functions in the computational-level decomposition is not determined by the functions that specify the steps in the algorithmic realization of the computation by states of the system. In this appendix I illustrate this point by a discussion of Marr and Poggio's (1979) computational-level theory of stereoscopic vision.

The two eyes receive slightly different views of the world due to their different positions in the head; stereopsis is a process whereby the difference between the two views is used by the visual system to determine the relative distances of objects outward from the eyes. The basis of stereopsis is the trigonometric relation between the distance of an object from the eyes and the angle formed by the two lines of sight to the object from each eye. In Figure 1, the two points

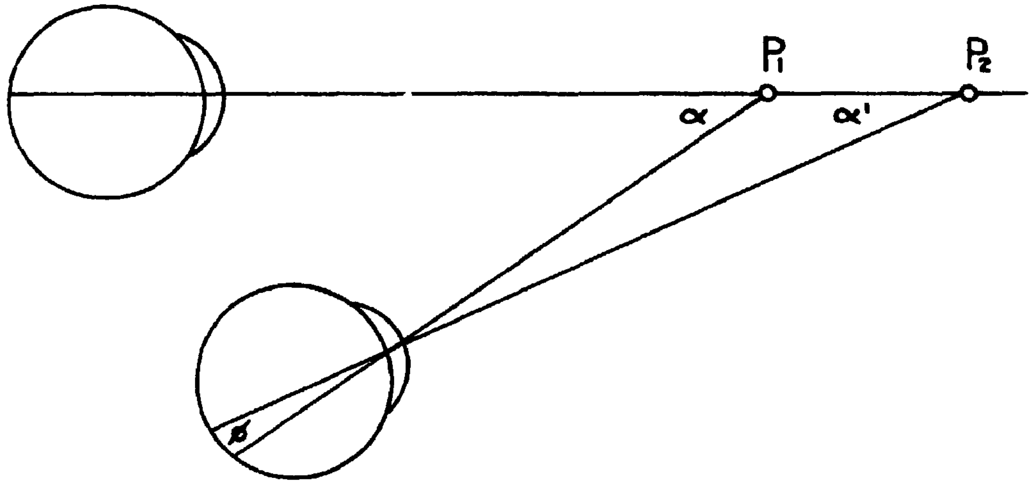


FIGURE ONE

P_1 and P_2 lie along a single line of sight from the left eye. As the illustration shows, the relative distances of points along a line of sight are a function of the differences in the values of α at each point. This difference is generally called <u>disparity</u>. Most importantly, disparity is a measure of the difference in angular distances between points in the right and left images. Thus in Figure 1, the difference between α and α' is equal to the angle ϕ.[1] Hence the relative distance of objects from the eye is recoverable from the relative distances between tokens in the two images.[2]

Since the slope of a plane surface is the integral of its distance, this principle can be extended to determine the orientation of plane surfaces in the distal scene. The relationship between disparity and distances along a line of sight illustrated in Figure 1, varies with change in the direction of the line of sight. But since this variation is regular, the orientation of a surface in front of the viewer is

---

1. To see this easily for yourself, close your right eye and place your two thumbs one behind the other in front of your left eye. Now close your left eye and open your right eye -- the distance between the two thumbs in the view of the right eye is disparity. Notice how the disparity varies with changes in relative distances of the two thumbs from the left eye.

2. Following Marr, I will use the term "token" to refer to any mechanically identifiable feature of the image. Although its use by Marr is tied to his particular theory of the primal sketch, its use here without such connotation maintains terminological continuity. As we have seen in Chapter 3, the term "feature" has its own theoretical associations.

recoverable from the rate at which disparity changes across the surface. Thus in Figure 2 the horizontal angle θ of the line L at P is a function of the rate at which α varies with change in the direction δ of L away from P. Accordingly for a plane surface we can define disparity in this sense as the instantaneous rate of change of α at a point on a surface. In Figure 3, disparity at P can be approximated from the difference in relative distances between l_1, l_2, and l_3 in the left image and r_1, r_2 and r_3 in the right image.

Stereoscopic vision, then, is asserted to be based on the relation between disparity and distance from the eye. Accordingly, part of the explanation of stereoscopic vision is the assertion that perceived depth is the result of the computation of a function from disparity. As a description of a relation between variables in the stimulus, this assertion makes no claims about how that relation is computed. Notice, however, that the assertion that depth perception is computed from disparity entails the claim that disparity is <u>available</u> to the visual system in the sense that the system must first somehow determine disparity values across the field of view. Since disparity is simply a difference between relative distances in the two images, it might appear that disparity is <u>directly available</u> to the system in the sense that there is a purely neurological description of a mechanism that compares the variation in distances across the two images.

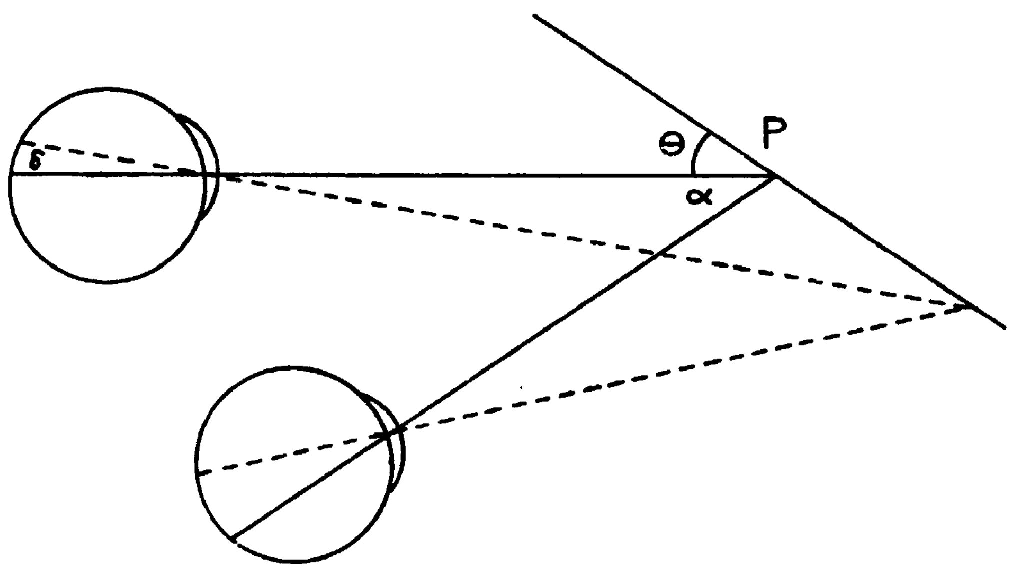


FIGURE TWO

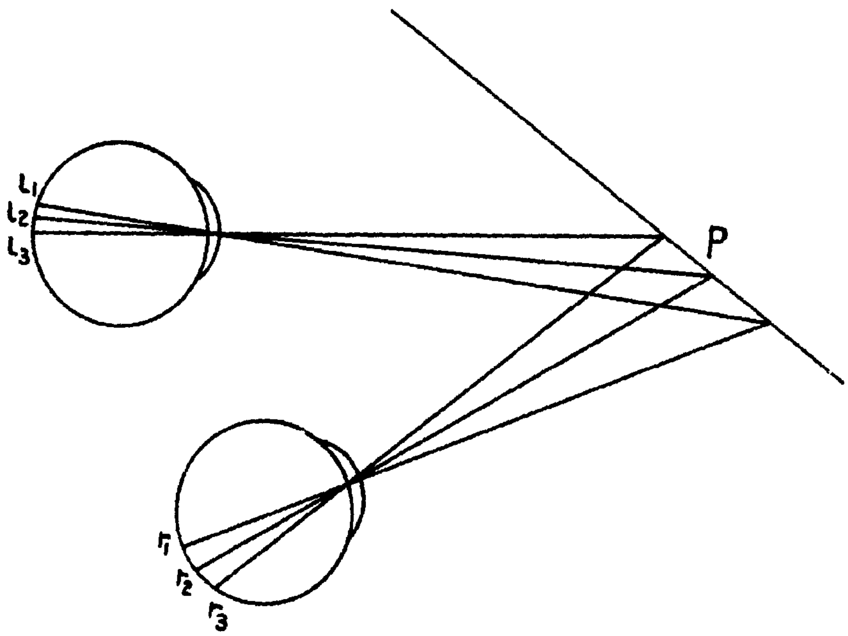


FIGURE THREE

However, the determination of disparity involves a distinct information-processing problem, the solution to which requires a computational-level description of a functional relation between stimulus variables. The problem, referred to as the matching problem, is to determine which pairs of image tokens should be matched in measuring positional discrepancy between the two images. Let us look more closely at the problem and at Marr and Poggio's (1976) solution to it.

It is only possible to determine disparity from relative distances in the right and left images if there is a way to match those tokens in the two images that are formed by the same points in the distal scene. Thus in Figure 3 there must be a way of matching each l_i with r_i. Notice that the correct matching -- that is, the matching that yields true disparity values across the field of view -- is one that satisfies a particular relation between the matched tokens of the images and points in the distal scene: A matching is correct if the matched pairs of image tokens are those formed by the same points in the distal scene. But this relation is not directly available to the system in the sense that there is a purely neurological description of how the system selects the pairs that stand in that relation. The mechanism cannot match tokens in virtue of the fact that they have the same distal origin. The problem, then, is to find another way of matching tokens that selects the same pairs but that refers only to some property P of the two images that is detectable by a mechanism in the system. So in addition to the neurological description of the mechanism that selects pairs on the basis of P there must also be an explanation of the fact that the selection yields the correct set.

The matching problem is one of a group of similar information-processing tasks referred to collectively as the correspondence problem. Just as in stereopsis, the stimulus variable from which distal properties are determined in the computation of shape from variation in two-dimensional projections is change in relative distances between image tokens with variation in viewing position; and there are other processes that utilize the same relation. In order to exploit such a variable in any of these operations the system must be able to determine which tokens in the image at different viewing positions are those formed by the same features in the distal stimulus. Notice that in all of its instances the correspondence problem is a matter of determining relations between tokens in the image and features of the distal scene. Thus whenever variation in spatial relations in the image is used to calculate spatial relations in the physical world, the system must somehow obtain prior information about the relation between the retinal image and the world.

Of course the matching problem in stereopsis is solvable if each object or marking in the distal scene is qualitatively distinct since, in such a case, there need only be a mechanism that matches identical tokens in each image. But this cannot be assumed to hold in the physical world; moreover, we know from an elegant set of experiments devised by Julesz that the visual system is able to solve the matching problem on the basis of the discrepancy in distances between tokens in each image. More precisely, from the difference in relative distances between tokens in a given region of the right and left images, the visual system determines a disparity value that is correct for images produced under normal conditions.

Marr and Poggio's (1979) solution to the matching problem is based on the fact that false matches produce descriptions of the distal scene that are inconsistent with certain global properties of the normal physical environment. Notice that each possible matching for a given region of the image determines a set of disparity values for that region, and thus also determines a set of surface orientation values for the region. Marr and Poggio's solution exploits this fact to eliminate those matches that yield descriptions of surfaces that do not occur in the normal environment. According to Marr and Poggio, the correct matching is described by three rules restricting the set of possible correct matches; each rule eliminates certain matches that cannot be correct in virtue of a specific global property of the physical world. The three rules are,

COMPATIBILITY: Only qualitatively similar tokens in the right and left images can be paired.

UNIQUENESS: Each token in one image can be paired with at most one token in the other image.

CONTINUITY: The disparity of matches varies smoothly; that is, disparities at adjacent points in the image will usually be nearly equal.

The first rule reflects the fact that only qualitatively similar image tokens could have been produced by the same distal feature. The second is based on the fact that only one point in the distal scene is visible along a single line of sight -- matching any token in either image with two tokens in the other image is presumably inconsistent with this fact. The third rule is suggested as a consequence of the fact that surfaces in the world are relatively smooth in comparison to area. Marr argues through an informal proof that a set of matches that satisfies these three rules is a correct solution to the matching problem.

Given the existence of the matching problem, stereopsis involves the computation of two distinct functions: a function F_1 that determines disparity values from distances in the two images, and a function F_2 that determines perceived depth from disparity. Yet we can also see that in one sense it is unnecessary to claim that disparity is computed by the system at all. For it is possible that the computation of depth from the positions of tokens in the two images is performed in a single step that has no *algorithmic* decomposition.

Given the function F_1 from distances across the two images to disparity, and the function F_2 from disparity to depth, there is a composite function $F_1\ F_2$ asserted to be computed by the system from distances between tokens in the right and left images to perceived depth. The point to notice, then, is that the *computation* of $F_1\ F_2$ need not have an algorithmic-level decomposition such that the system computes the sequence $\langle F_1, F_2 \rangle$. That is, the algorithmic-level explanation of the computation of $F_1\ F_2$ may not involve a mapping from physical states of the system to the values in the range of F_1; the interpretation function I may simply be a mapping from the computational states of the system to the union of the domain of F_1 and the range of F_2.

But while this is true, we can also see that the *computational-level* description of stereopsis must decompose the operation into two functions to explain the dependence of the success of depth perception on contingent facts about the physical world. Each rule in Marr and Poggio's solution to the matching problem is based on the assumption of a specific global property -- a natural constraint -- of the world. As we have seen, part of the explanation of stereopsis must be an explanation of how the properties of the image used by the system to solve the matching problem determines the correct set of pairs. Hence the theory must include an explanation of how the solution to the matching problem depends for its success on the natural constraints. But this is possible only if the description of the relation between image and percept is decomposed into two operations: (1) calculation of disparity from relative distances in the two images, which involves matching pairs of tokens in the images, and (2) calculation of depth or surface orientation from disparity. Hence the perception of depth in stereo vision must be seen as the product of two operations independently of whether or not the realization of these operations has an algorithmic-level decomposition.

Of course, the point does not depend in any way on the specific content of Marr and Poggio's solution to the matching problem. There must always be some description of the pairs over which disparity is determined; hence if Marr and Poggio's solution is not experimentally confirmed, there is some other function that describes the pairs from which disparity values are taken. (In fact there is evidence that the uniqueness constraint is not obeyed in human vision. See Weinshall 1989.) And there must be an explanation of how this matching is veridical under normal conditions. Whatever the solution to the matching problem turns out to be, then, a computational-level theory of stereopsis must decompose the operation into solutions to two distinct information-processing tasks, regardless of the number of steps in the algorithmic-decomposition of the operation.

APPENDIX THREE

NEUROPHYSIOLOGY AND IMAGE INTERPRETATION

In Chapter 3 I argued that neurophysiological theories of visual processes are explanatorily incomplete. Understanding vision requires a computational-level theory which, as we saw in Chapter 1, is autonomous with respect to the algorithmic level of description, and hence also with respect to the physical level. The role of the computational level is to describe the relations between stimulus variables upon which the veridicality of visual representations depends, and to do so in a way that specifies precisely how these relations can be mechanically computed. This task, we saw, is logically prior to, and unaffected by, the question how the relations are actually computed by the system. In this discussion I described Marr's criticism of the feature-detector theory of vision which, according to Marr, is unsuccessful because it ignored computational-level theories in favour of neurophysiological investigation. In this appendix I illustrate this point by describing how Marr and Hildreth's (1980) theory of edge detection provides a clear computational-level interpretation of neurophysiological observations, and thus <u>explains</u> these observations in a way that is impossible without such an interpretation.

Recall from Chapter 3 that the feature-detector strategy is based on the fact that the responses of single cells in the neuronal structure of the visual system signal the presence of patterns of light intensities in specific regions of the retinal image. It was experimentally discovered that the behaviour of higher-level cells in the visual system is determined by the individual responses of collections of receptor cells in different regions of the retina. Since the firing of a receptor cell may either inhibit or excite the firing of a higher-level cell, the latter can be thought of as responding selectively to specific variations in light intensity in the grey level image.

The idea here is quite simple. In order to detect changes in light intensity in the grey level image it is necessary only to compare the intensity at one location with that at another. The simplest way to do this is to subtract one from the other; the difference will be a measure of the intensity gradient across the spatial interval separating the two points. A simple mechanical way of doing this is to weight the two values by -1 and +1 respectively and take the sum of the two results. Thus the excitatory and inhibitory cells may be thought of as weighting the light intensity values at different points by a negative and positive factor; the higher-level cell will fire if the difference in value across the region is within a certain interval indicating a gradient within a given range. The initial work in this domain was

Kuffler's (1953) discovery of the centre-surround structure of the receptive field of retinal ganglion cells[1]: Receptor cells in the central region of the receptive field provide excitatory input to the ganglion cell while those in the surrounding region provide inhibitory input (and vice versa).

From these discoveries it is argued that higher-level cells in the system perform complex information-processing tasks the results of which carry specific information about the stimulus. The inference is supported by a number of subsequent studies. For example, Lettvin et. al. (1959) conclude that signals of nerve fibres in the retinal ganglion of the frog indicate four specific features of the image independently of general illumination: local sharp edges, dark objects with convex edges, moving edges, and sudden reductions in illumination. And Hubel and Wiesel's extensive study of the receptive fields of cells in the lateral geniculate nucleus and in the visual cortex suggest that cortical cells are sensitive to bars and edges at different orientations in the image.

According to early theories, these results indicate that visual processes detect specific <u>features</u> of the image, collections of which are correlated with properties of the distal stimulus. This idea suggests that understanding vision can be accomplished by determining,

---

[1]. The receptive field of a cell is simply the area of the retinal image to which it is responsive.

through direct neurophysiological investigation, the features to which cells in different levels of neurological structure are responsive. Yet it proved difficult to extend the initial results. Part of the problem is simply the difficulty involved in specifying the image features to which higher-level cells are responding. Thus Marr argues that the range of stimuli to which cells as described by Hubel and Wiesel would respond is not specific enough to be interpreted as indicating the presence of any particular feature in the image; for example, those with bar-shaped receptive fields would respond equally to the presence of a bright edge or a dim bar.

But this difficulty is symptomatic of a deeper problem. The method of theory construction suggested by the feature-detector idea is to describe the extraction of information from the image by describing the behaviour of neural units. But as we have seen, understanding the role played by perceptual mechanisms is not obtained solely through descriptions of their behaviour; behaviour must be explained by specifying the information-processing problem for which it provides solutions. Even if it were possible to specify precisely the image features to which a cell is responsive, such knowledge can be used to explain the working of the system only if we also know what information about the stimulus such a feature is <u>interpreted</u> as carrying. If, as Marr contends, visual processes exploit relations between higher-order variables in the stimulus that result from natural constraints, providing interpretations of neural activity will require a precise description of these relations, and of the natural constraints upon which they rest. To see this clearly, let us look at Marr and Hildreth's (1980) interpretation of the results obtained in the feature detector studies.

Recall from Appendix 2 that the correspondence problem -- for example, the matching problem in stereopsis -- is the question how to match items in the image at different views; but what are the items to be matched? Any solution to the correspondence problem that does not employ specific information about the distal scene will depend on some set of global natural constraints to reduce the set of possible matches to one. But such solutions are possible only if the items that are matched reflect properties of the distal stimulus, such as shadows and edges. So solving the correspondence problem presupposes a means of locating tokens of the image that are, as Marr puts it, "physically meaningful". The work begun under the feature-detector program has a precise formulation as a computational-level problem: What operations on the image will specify the locations and descriptions of items in the geometrical structure of the image that reflect properties of the distal scene? This is the task of generating the primal sketch from the grey level array. The reasoning of Marr and Hildreth in describing the raw primal sketch is roughly as follows.

A glance at any two-dimensional, black and white image will show that properties of physical surfaces, such as shadows and edges, are often reflected in the image by changes in the level of light intensity. So the first part of the task might be to find the locations of changes in light intensity in the grey level image; this possibility receives support from the structure of receptive fields. Moreover, there is a natural way to locate these changes in the image. Changes in light intensity from dark to light, or vice versa, are indicated by critical points, i.e., points in the image at which variation in intensity gradient in a particular direction changes from an increasing to a decreasing gradient. If we think of light intensities <u>in a single given direction</u> across the grey level image as values of a single-valued differentiable function, these points are indicated by peaks in the first-derivative of the function or by zero-crossings in the second-derivative. The task of locating intensity changes in the image is thus equivalent to locating critical points in the grey level image – finding the peaks the first derivative or zero-crossings in the second-derivative -- in any direction across the image.

However, according to Marr and Hildreth, two features of the physical world impose conflicting constraints on the task of accurately locating and describing changes in light intensity that reflect properties of the world:

(1) The changes in light intensity that the operators are required to detect occur at a range of scales. Evidently a range of operators is required, where each operator responds effectively for a narrow range of scales. For if intensity levels are compared over a very small interval, the difference in intensity levels for very slow changes in intensity over a large area may be too small to be detected by an operator. And, on the other hand, if the intensity levels are compared over a large interval, small sudden changes in intensity will be missed altogether.

(2) The tokens that the operators are required to detect are "spatially localized" at their own scale, in the sense that they do not consist of repeating patterns like ripples on a beach.

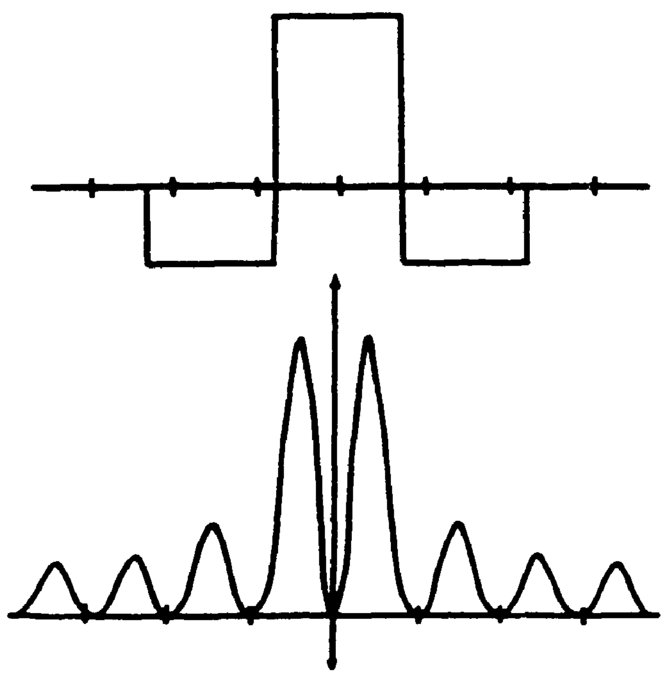
The difficulty that Marr and Hildreth point to is that we cannot simultaneously minimize the range of scales to which the operator is sensitive and maximize the accuracy with which critical points are detected. There is now a clear formulation of a computational-level problem: Find an operator which, when applied to the image, achieves the best possible results in both range of intensity changes and spatial localization. To see how Marr and Hildreth's solution to this problem provides a computational-level interpretation of observations of neural behaviour, let us look at the problem and their solution in more detail.[2]

First notice that locating critical points in the image at a given scale is equivalent to finding the critical points in the image after it has been blurred by a smoothing function. The value of a smoothing function for a point in the image is an average of intensity levels in the surrounding region; the larger the region over which intensity levels are averaged, the greater the blurring of the original image. Such a smoothing function has the effect of wiping out intensity changes below a certain scale. We can combine such a smoothing function with an operator that produces an approximation to the second-derivative in a single direction by an operator whose weighting function has the graph illustrated in Figure 4(a). The horizontal axis of the graph represents a set of points along the image in one direction; the value of the graph at each point along the horizontal axis indicates the weighting factor applied to the intensity level at that point in the image. The value of the operator at a point P in the image is the sum of the weighted values. We can think of the neuronal structure of the system as having a collection of such

---

[2]. The problem described in the following paragraphs has a formal similarity to the uncertainty principle in quantum mechanics although it is not a quantum mechanical effect. The problem discussed below that arises in d.scribing an appropriate operator to detect intensity changes is a consequence of the fact that the output graph of such an operator is its Fourier transform.

(a)



(b)

FIGURE FOUR

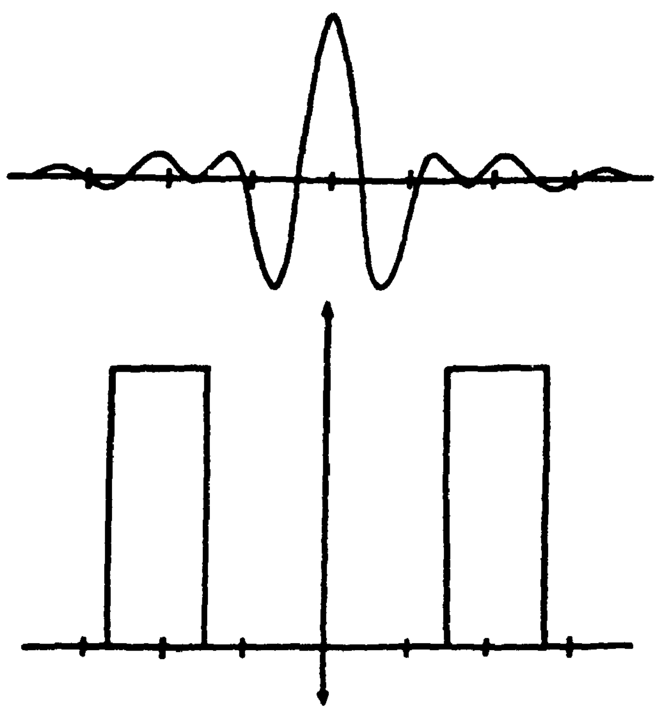
operators in different directions across the entire retina.

Now if, in engineering terms, we think of light intensities across the image in a given direction as a signal varying in space rather than in time, then a smoothing function can be thought of as a filter that allows only a specific range of "spatial frequencies".[3] In these terms, the operator depicted in 4(a) is a filter whose output indicates zero-crossings in the second derivative at certain spatial frequencies but not others. We can represent the response of such an operator by another graph where the horizontal axis represents spatial frequencies, and where the values of the graph represent the strength of output of the operator at each spatial frequency. Then an operator described by the graph in 4(a) will have a response graph of the shape illustrated in 4(b). Clearly such an operator is not ideal in the sense that it responds to a narrow range of spatial frequencies, and in fact it will miss some critical points in the image. On the other hand, if we define an operator that responds only to a narrow range of spatial frequencies as illustrated by Figure 5(b), the weighting function of such an operator has a graph like that illustrated in Figure 5(a). This second operator is not suitable since the side lobes in the

---

[3]. The spatial frequencies of a signal are the components of its Fourier transform, where the Fourier transform is a representation of the signal as the sum of a set of sine and cosine waves.

(a)



(b)

FIGURE FIVE

weighting function will produce echoes in the form of responses in the output profile that have no correlates in the image.

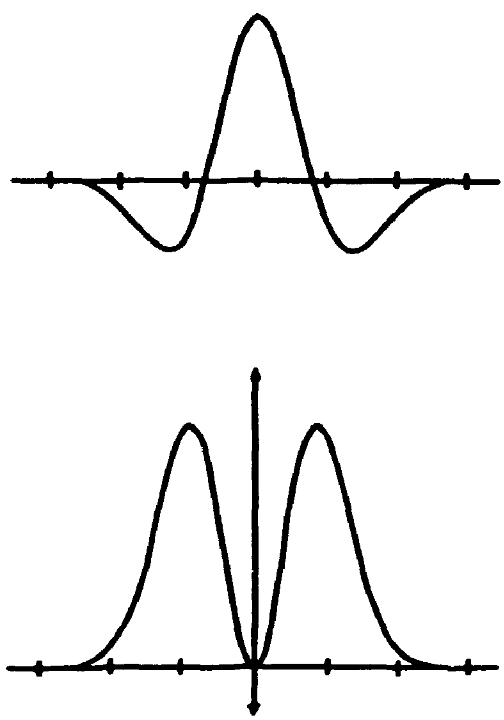
Although the conflict here arises from mathematical facts, Marr and Hildreth point out that the conflicting requirements on operators are dictated by the structure of the physical world. Marr and Hildreth's solution is to blur the image by applying a weighting function with the form of a Gaussian or normal distribution. Combining such a weighting function with an operator that produces an approximation to the second-derivative yields an operator with a graph of the form depicted in 6(a); the output profile of this operator is that depicted in 6(b).[4]

The operators we have been discussing apply only in a single direction at a time so that we have been considering only the detection of intensity changes in a single-valued function in one direction. But the grey level image is a two-valued function where the visual system will need to detect critical points in any direction. We might solve this problem by using a collection of operators at different orientations, a solution first tried by Marr (1976). However, Marr and Hildreth argue that the isotropic second-order Laplacian operator $\nabla 2$ will detect intensity changes in any direction, given certain assumptions about the image. Hence they conclude that the ideal

---

[4]. The basis of this solution is the fact that only a Gaussian distribution is its own Fourier transform.

(a)



(b)

FIGURE SIX

operator is the Laplacian operator applied to an image blurred by a Gaussian weighting function. These two operations can be combined in a single operator ∇2G, the weighting function of which has the "Mexican hat" form of the profile depicted in Figure 6(a) rotated around its centre.

But notice that, as a computational-level result, the ∇2G operator describes only the relation between the image and the perceived intensity changes, not the computation of this relation by the system. Marr and Hildreth point out that the ∇2G operator is closely approximated by a function obtained by taking the difference between two Gaussian distributions, -- a function that closely resembles the form of the centre-surround receptive fields of cells in the early visual system. There is, then, neurological evidence that the system determines the result of applying the ∇2G operator to the image by taking the difference between the values obtained from two Gaussian weighting functions. Here, then, there is a clear distinction between the computational-level demonstration that the computation of a particular function solves an information-processing problem, and the description of how that function is computed by the system. Thus the computational-level account provides a clear <u>interpretation</u> of the behaviour of the neural units that is lacking in the feature-detector approach. Marr's defence of the computational level is precisely that it provides a clearly defined set of information-processing tasks performed by the visual system to serve as a basis for computational modelling.

APPENDIX FOUR

THE ROLE OF EMPIRICISM IN HELMHOLTZ'S THEORY OF VISION

The purpose of this appendix is to suggest a reading of Helmholtz's empiricism in perception theory. On this reading empiricism is the only possible methodological principle that will yield explanations of perceptual phenomena. Properly understood, the function of empiricism in Helmholtz's explanations of vision is similar to the role of Marr's computational level of description.

My argument will rest on a description of a particular theory in The Treatise on Physiological Optics. The theory is selected to make a specific point. Helmholtz's empiricism has been criticised for its mentalist features, i.e., for its appeal to unconscious inference in perception. But much of this criticism is based, I believe, on incorrect understandings of the methodological and explanatory roles his empiricism plays on his theories.

Gibson argues that Helmholtz's empiricism, and his mentalist descriptions of perceptual processes, result from the fact that Helmholtz attempts to find the image correlates of properties of the physical world in a momentary, static image. Thus, according to Gibson, Helmholtz's empiricism is based on the underdetermination of three-dimensional properties of the world by momentary static cues.

But in the particular example that I will discuss, there is no such underdetermination. Rather, Helmholtz's theory is intended to explain how properties of a monocular image -- viz., straight lines -- are detected by visual processes when they are not <u>directly detectable</u> in the sense described in Section 4 of Chapter 1, that is, when an explanation of the response of the visual system cannot be given just in terms of the stimulus property and neural processes of the system.

Meyering (1989) argues that Helmholtz's mentalism derives from a methodological conservatism: since <u>inference</u> is a familiar and well-established psychological process, it is preferable to seek explanations of mental abilities in terms of unconscious inference when there are no other well-understood explanations. But I argue that Helmholtz's empiricism, and thereby his appeal to unconscious inference also, is not based on assumptions about the mechanisms by which the solutions to perceptual problems are implemented. Rather, like Marr's computational level of description, it has a methodological role in uncovering the relations between stimulus properties that provide the basis for perceptual abilities. Helmholtz's solutions to visual tasks thus provide a basis for <u>computational</u> theories just as readily as they do for theories of unconscious inference.

Let us look first at how Helmholtz understands his empiricism. Certainly part of Helmholtz's empiricism is the assertion that the correlations between properties of the image and properties of the world which are used in visual perception must be learned through experience. But a better understanding of its methodological role is gained by taking empiricism to be the more modest claim that the psychologically primitive contents of the mind in perception specify only states of the sensory system and not the spatial locations of features of the world.

It follows that the perception of direction and distance are the work of the intellect. This interpretation is suggested in the third volume of the Physiological Optics where he gives as the fundamental principle of empiricism the "theory of signs".

> The sensations of the senses are tokens for our consciousness, it being left to our intelligence to learn how to comprehend their meaning.
> (1867, Vol.3: p. 533)

In Helmholtz's view, the theory of signs is an extension of Müller's doctrine of specific nerve energies. For example, in the opening section of the second volume of the Physiological Optics he says,

> Whether the sun's rays will be perceived as light or heat, is simply a question of whether they are perceived by the optic nerve or by the cutaneous nerves. But whether they will be perceived as light that is red or blue, and dim or bright, or as heat that is mild or intense, depends both on the nature of the radiation and on the condition of the nerve. The quality of the sensation is thus in no way identical with the quality of the object by which it is aroused. Physically, it is merely an effect of the external quality on a particular nervous apparatus. The quality of the sensation is, so to speak, merely a symbol for our imagination, a sort of earmark of objective quality.

The first two sentences of this passage are Müller's doctrine; the last sentence is Helmholtz's theory of signs.

Intuition theory, to which Helmholtz's empiricism is opposed, is described in the following passage from the third volume.

> The cardinal fact about [all intuition theories] is that the localization of the impressions in the field of view is derived through some innate contrivance, and either the mind is supposed to have some direct knowledge of the dimensions of the retina, or it is assumed that, as the result of the stimulation of definite nerve fibres, certain apperceptions of space arise by virtue of an innate mechanism that cannot be further defined. (1867, Vol. 3: 541f)

So empiricism is contrasted with the view that the perception of spatial location is either a matter of direct knowledge or the immediate result of nervous activity.

Why does Helmholtz believe that spatial perception requires the operation of the intellect? He gives several reasons in various places, but in his discussions of specific perceptual phenomena he almost always claims that intuition theory is unable to account for these phenomena without ad hoc revisions. So his constant view is that empiricism is the only methodological principle that will yield explanations of perception.

My example for this claim is Helmholtz's discussion of the relation between ocular movements and apparent size and direction in the visual field. To this end let me describe two laws of ocular movement discussed by Helmholtz.

Consider the eye first in its natural, or primary, position looking straight ahead. Any change in the direction of the line of sight can be described by the elevation and azimuth of the secondary position as measured from the primary position. However, as well as changing the direction of the line of sight the eyeball may also rotate around that line. Let us call rotation of the eyeball around the line of sight, the torsional rotation.

The first point to note is that there must be some torsion in the movement of the eye. The eyeball will move to a new position with no torsion only if it rotates about an axis that lies in the plane perpendicular to the primary line of sight, roughly the plane formed by the two eyes and the chin. But any movement from a secondary position that does not return the eye to the primary position involves some degree of torsion. Using only rotations about an axis in this plane, the eye can move from one secondary position to another only by first returning to the primary position, which we know is not the case.

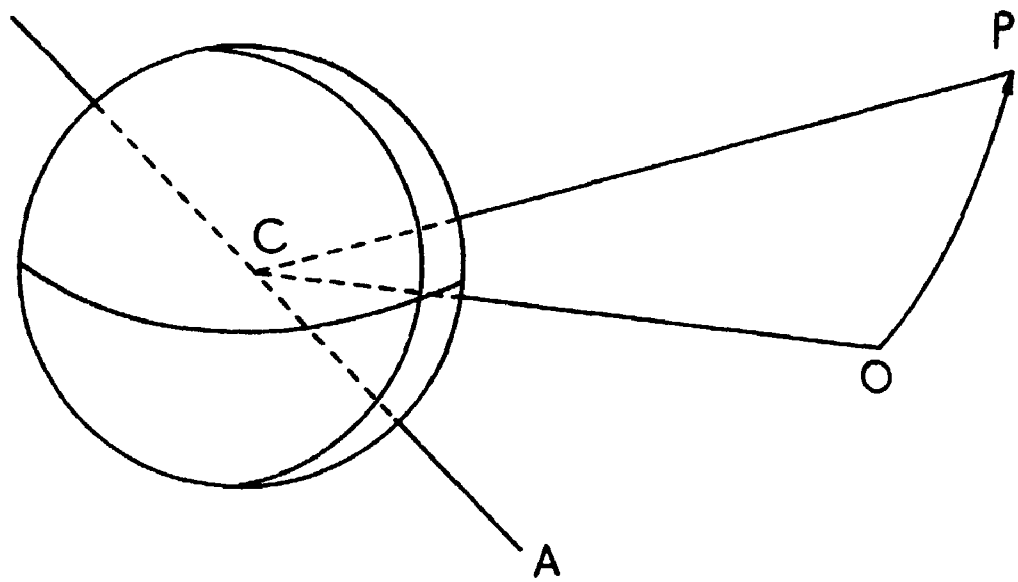
The two laws of interest for us are Donder's law and Listing's law. Donder's law is that the eye always adopts the same amount of torsion for every line of sight. Listing's law tells us what the amount of torsion is for each line of sight. Listing's law is as illustrated in Figure 7.

> The amount of torsional rotation of the eyeball when the line of sight moves from the primary position to a secondary position is the same as if the eye rotated around an axis perpendicular to both the primary and secondary lines of sight.

It follows from Listing's law that there is no torsional rotation when the eye is simply raised or lowered, or moved to one side, from the primary position since the plane of rotation in these cases is perpendicular to the primary line of sight. But there is <u>clockwise</u> rotation when the eye moves toward the upper left or lower right, and <u>counterclockwise</u> rotation when it moves toward the lower left or upper right.

The importance of these laws for estimations of the visual field is that a line that appears horizontal when viewed directly from the primary position will be tilted from the horizontal whenever there is torsional rotation of the eye. Thus a horizontal line will appear tilted when viewed in any of the four corners of the field of view. It will appear tilted counterclockwise when viewed in the upper left or lower right extremes of the field of view, and will appear tilted clockwise when viewed in the lower left or upper right extremes.

There is one last point to make with respect to eye movement. Let us call the arc traced in the field of view by the line of vision when the eye moves in accordance with Listing's law, a <u>direction line</u>. We can show that every direction line, if continued, meets the point in the visual field directly behind the line of sight



Listing's Law

The amount of tortional rotation of the eyeball when the point of fixation moves from O to P is the same as if the eyeball rotated around an axis passing through the centre of rotation C and perpendicular to the plane OCP.

FIGURE SEVEN

when the eye is in the primary position. When the eye is raised or lowered, or moved to one side, from the primary position, the direction lines will be great circles in the field of view; otherwise they will be circles of smaller diameter.

So much for eye movement; let us now look at determinations of size and direction in the field of vision. In particular, I want to look at an illusion that occurs in the perception of straight lines in the two-dimensional monocular field of view.

We can consider the visual field without depth to be a sphere with its centre at the centre of rotation of the eye. If we ignore the difference between the centre of rotation of the eye and the point of intersection of the rays of light in the eye, we can imagine the eyeball and its field of view as concentric spheres. As the eye moves in its socket, a point in the field of view will be projected onto different points on the retina.

Since the visual field without depth is a sphere, straight lines in the visual field are great circles. And since the visual field and the eye can be considered concentric spheres, great circles in the visual field are projected onto great circles on the sphere of the eye. Hence, we would expect that lines that appear straight to the eye are great circles in the visual field. <u>But this is not the case</u>. Lines that appear straight are the direction lines along which the eye moves according to Listing's law. As we have seen, these will be great circles only in the case of movement up or down, or to one side, from the primary position. In the periphery of vision, lines that are great circles in the visual field will appear concave inward. Conversely, lines in the visual field that are convex inward will appear straight. Thus the lines of the checkerboard in Figure 8 will appear straight when viewed close to with one eye. We can also notice this illusion by looking down at a point on a table and arranging three pieces of paper in what appears to be a straight line in the extreme periphery of vision; the papers are in fact placed in a convex arc. As Helmholtz points out, the effect is as if the rays of light entering the eye intersect at the back of the eye rather than near the centre of the eye.

The question is, what is the reason for this particular illusion? Helmholtz contends here, as he does in every case like this, that intuitionists can give no explanation for this phenomenon. I believe his meaning is that, for the nativist, since the order and distance of points in the stimulation of the eye is precisely the same as that of points in the visual field, all that is needed for the determination of distance and direction in the visual field is for this information to be carried to the sensorium.

On this view, there is no particular reason why the apparent direction of lines should be different than the actual lines. Once the

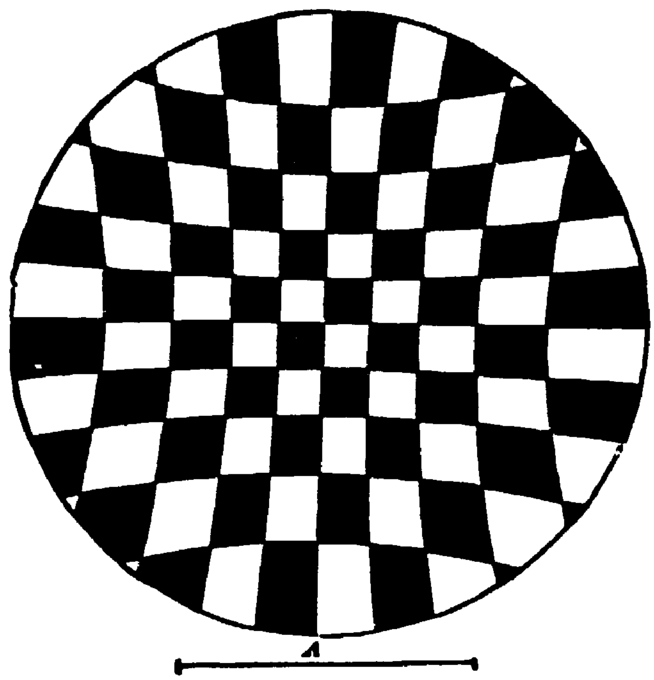


FIGURE EIGHT

illusion is noticed, it is always possible for the intuitionist to describe the nervous mechanism accordingly; but this will be an ad hoc manoeuvre rather than a genuine explanation.

Let us look at the illusion from what I think is Helmholtz's perspective. I have rearranged the order in which Helmholtz makes the points I describe, but I do not think this alters his point.

One of the things we wish to explain is our ability to estimate with reasonable accuracy when a set of points lies along a straight line. Given the correspondence between points in the field of view and points on the eye, this is possible if we are able to determine great circles in the order of stimulation on the retina. The problem is that we need some way to determine these great circles locally. It is not directly apparent in vision that a line lies on a circle whose centre coincides with the centre of the eye. In the terms developed in the discussion of Marr's work, great circles are not directly detectable by the visual system.

The solution Helmholtz suggests is based on the fact that straight lines in the world are such that every part is congruent with every other part so that they can be moved along themselves. The only other lines that have this property are circles, but circles are congruent only when superposed in a certain way. (1867, Vol 3: p. 176) Accordingly, we can determine straight lines in vision as those that can be shifted along themselves. But given the nature of eye movements, the lines that have this property in fact are not great circles but the direction-lines. Thus the illusion above is explained if we assume that this property of lines is used to determine straight lines. He says,

> Now in the field of fixation there is only one species of lines which require only a direct act of sensation for us to tell whether they can be shifted along themselves and are therefore congruent with themselves all over. As shown by the preceding investigation, on the assumption of Listing's law, these lines are the direction-circles. It is true that there may be other circles in the field of fixation which must be admitted this same property, but we cannot prove it except by measurements and deductions, not by a direct act of sensation.
>
> In the field of view, instead of having a ruler that can be shifted, we have the central place where vision is most distinct ... We shift the gaze along this line, thereby shifting the line itself and indicating to ourselves the continuation of this direction. (1867, Vol 3: pp. 176f)

The point is that even when information about the world is available in the image projected onto the eye, there must still be an explanation of how it is extracted.

In the case described above, this is possible only if the visual system assumes, implicitly or explicitly, that straight lines are those whose parts are everywhere congruent. Because of the restrictions on eye movement, this assumption is false; hence the illusion. But if the eye moves according to Listing's law there are lines in vision that are congruent with themselves everywhere. And as Helmholtz demonstrates, movements according to Listing's law minimize the errors caused by torsional rotation so that the illusion under this principle is the least possible.

Here is what I think this example of Helmholtz's explanations of perception illustrates. The task of perception theory is to explain perceptual abilities -- we need to explain how we arrive at a knowledge of spatial relations from retinal stimulation. This ability involves extracting the information from available properties of the stimulus array. The question in the case of each perceptual ability is, what assumptions about relations between stimulus properties makes this ability possible? The theory is confirmed when perceptual illusions are successfully predicted by the assumptions that connect retinal properties with spatial relations. Intuition theory skirts this issue, and thus cannot predict illusions or veridical perceptions from principles established on grounds independent of the phenomena to be explained. It can account for perceptual phenomena only on the basis of ad hoc principles designed solely to accommodate the observed regularities. So, according Helmholtz, empiricism is the only genuinely explanatory methodology.

Helmholtz's explanation of the perception of straight lines in the visual field shows how specific perceptual illusions can be predicted on the hypothesis that a certain relation between stimulus properties is exploited in monocular perception. On Helmholtz's view, relations between stimulus properties can be exploited in this way only by unconscious inference based on experience. But the theory itself is independent of this view. The explanatory value of the theory lies in the framework it suggests for uncovering the relations in the stimulus that underlie perceptual abilities, and not in the descriptions of mechanisms by which these abilities are realized.

BIBLIOGRAPHY

Barlow, H.P. (1972). "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?" Perception 1: 371-394.

Burge, T. (1986). "Individualism and Psychology." The Philosophical Review 95: 3-45.

Chomsky, N. (1959). "A Review of Skinner's Verbal Behaviour." Language 35: 26-58.

Churchland, P.S. (1986). Neurophilosophy: Toward a Unified Science of Mind-Brain, Cambridge, Mass.: M.I.T. Press.

Churchland, P.M. (1984). Matter and Consciousness, Cambridge, Mass.: M.I.T. Press.

Davidson, D. (1970). "Mental Events." In Essays on Actions and Events, Oxford: Oxford University Press.

Demopoulos, W. (1987). "On Some Fundamental Distinctions of Computationalism." Synthese 70: 79-96.

Fodor, J.A. (1974). "Special Sciences." Synthese 28: 77-115. Reprinted in Fodor (1975) and(1981).

Fodor, J.A. (1975). The Language of Thought, Cambridge, Mass.: Harvard University Press.

Fodor, J.A. (1980). "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." Behavioural and Brain Sciences 3. Reprinted in Fodor (1981).

Fodor, J.A. (1981) Representations: Philosophical Essays on the Foundations of Cognitive Science, Cambridge: M.I.T Press.

Fodor, J.A. (1983). The Modularity of Mind, Cambridge, Mass.: M.I.T. Press.

Fodor, J.A. (1986). "Off The Slippery Slope or Why Paramecia Don't Have Mental Representations." In P.A. French et. al. (eds.) Studies in the Philosophy of Mind, Midwest Studies in Philosophy, Vol. X, Minneapolis: University of Minnesota Press.

Fodor, J.A. (1987). Psychosemantics, Cambridge, Mass.: M.I.T. Press.

Fodor, J.A. and Block, N. (1973). "Cognitivism and the Analog/Digital Distinction." privately circulated.

Fodor, J.A. and Pylyshyn, Z.W. (1981). "How Direct Is Visual Perception? Some Reflections on Gibson's 'Ecological Approach'." Cognition 9: 139-96.

Gibson, J.J. (1955) The Perception of the Visual World, Boston: Houghton Mifflin.

Gibson, J.J. (1966) The Sense Considered as Perceptual Systems, Boston: Houghton Mifflin.

Gibson, J.J. (1979). The Ecological Approach to Visual Perception, Boston: Houghton Mifflin.

Gould, S.J. (1980). The Panda's Thumb, New York: Norton.

Helmholtz, H. (1867). Handbuch der Physiologischen Optik, Leipzig: Voss. Translated as Treatise on Physiological Optics, New York: Optical Society of America.

Kitcher, P. (1988). "Marr's Computational Theory of Vision." Philosophy of Science 55: 1-24.

Kuffler, S.W. (1953), "Discharge Patterns and Functional Organization of Mammalian Retina." Journal of Neurophysiology 16: 37-68.

Lettvin, J.Y., Matturana, R.R., McCulloch, W.S., and Pitts, W.H. (1959). "What the Frog's Eye Tells the Frog's Brain." Proceedings of the IRE 47: 1940-51.

Lycan, W.G. (1981). "Form, Funtion and Feel." The Journal of Philosophy 78: 24-50.

Matthews, R.J. (1986). "How Not to Mett the Eliminativist Challenge: A Reply to Fodor." In P.A. French et. al. (eds.) Studies in the Philosophy of Mind, Midwest Studies in Philosophy, Vol. X, Minneapolis: University of Minnesota Press.

Marr, D. (1976). "Early Processing of Visual Information." Phil. Trans. R. Soc. Lond. B 275: 483-524.

Marr, D. (1982). Vision, New York: W.H. Freeman.

Marr, D. and Hildreth, E. (1980). "Theory of Edge Detection." Proc. R. Soc. Lond. B 207: 187-217.

Marr D. and Poggio T. (1979) "A Computational Theory of Human Stereo Vision." Proc. R. Soc. Lond. B 204: 301-328.

Meyering, T.C. (1989). Historical Roots of Cognitive Science: The Rise of a Cognitive Theory of Perception from Antiquity to the Nineteenth Century, Dordrecht: Kluver.

Putnam, H. (1967). "The Nature of Mental States." In Mind, Language and Reality: Philosophical Papers, Volume 2, Cambridge: Cambridge University Press.

Putnam, H. (1975). "The Meaning of 'Meaning'." In Mind, Language and Reality: Philosophical Papers Volume II, Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1984). Computation and Cognition: Toward a Foundation for Cognitive Science, Cambridge, Mass.: M.I.T. Press.

Ramachandran, V.S. (1985). "The Neuruobiology of Perception." Perception 14: 97-103.

Stabler, E.P. (1983). "How Are Grammars Represented?" Behavioural and Brain Sciences 6: 391-421.

Stabler, E.P. (1987). "Kripke on Functionalism and Automata." Synthese 70: 1-22.

Ullman, S. (1979). The Interpretation of Visual Motion, Cambridge, Mass.: M.I.T. Press.

Ullman, S. (1980). "Against Direct Perception." Behavioural and Brain Sciences 3: 373-415.

Weinshall, D. (1989). "Perception of Multiple Transparent Planes in Stereo Vision." Nature 341: 737-739.

Wilson, M. (1985). "What Is This Thing Called 'Pain'? -- The Philosophy of Science Behind the Contemporary Debate." Pacific Philosophical Quarterly 66: 227-267.