VISUAL PERCEPTION

An Information-based Approach to Understanding Biological and Artificial Vision

by

Noel A. Murphy B.A.(Mod)

Submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Supervised by Prof. Charles McCorkell

School of Electronic Engineering
Dublin City University
Dublin 9 Ireland

September 1992

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work

Signed: Noel Murphy Date: 30/Sept/1992

Dedication

To Karina,

for all her love, support, patience and understanding,
especially over the last three years.

To Mum and Dad,

Brigid and Michael Murphy, who have given me enormous support right through my education and who have never seen me lacking for anything.

Acknowledgements

I would like to thank my supervisor, Prof. Charles McCorkell, a man of great patience and forbearance for giving me the opportunity to carry out this work at a critical stage of my career, for giving me the opportunity to discover that lecturing (in moderate quantities) and research (in immoderate quantities) were two things that I would find enormously fulfilling, and for the constant help, support and guidance throughout the long road to this dissertation.

I have had contact with many fellow students, and students of my own, in my years in the NIHE and now DCU. Particular names that spring to mind are Brian Buggy who started the same week as me and was a great friend and colleague throughout, Gearoid Mooney, Mark McDonnell, Jim Gunning and the many fourth-year B.Eng project students and who helped in various aspects of the project. For their interest, support and encouragement I am very greatful.

The staff in the School of Electronic Engineering, DCU have been very supportive and cooperative in all the time I have worked with them, and for this I would like to heartily
thank them. In particular I owe an enormous debt to Barry McMullin who introduced me to much of the literature which I have found most valuable, who kept me in touch with the outside world, who was always available to discuss and debate ideas which at the time seemed to have little place in an engineering department, and for the effort he put into organising the workshop on *Autopoiesis & Perception* held in DCU, August 1992, while he was preparing his own Ph.D. thesis. Also I am very greatful for the support and encouragement of Paul Whelan and Niall Byrne – Paul for his organisational contributions to the Advanced Vision Systems Group and the vision lab in DCU over the last two years, Niall for assisting in the preparation of various papers and presentations and for proofreading parts of this dissertation. John Whelan and his cohort of merry technicians were always there to sort out our problems, even the equipment ones, and their work and assistance is much appreciated.

I would like to thank Dr Eugene Kennedy, and Dr Michael Clancy who supported the initial setting up of this project, and the National Board of Science and Technology (NBST), now called EOLAS, who provided funding for the first two years work.

Table of Contents

Abstract

Visual Perception:
An information-based approach to understanding biological and artificial vision

Noel Murphy, School of Electronic Engineering, Dublin City University.
Ph.D. Thesis, September 1992.

The central issues of this dissertation are (a) what should we be doing – what problems should we be trying to solve – in order to build computer vision systems, and (b) what relevance biological vision has to the solution of these problems. The approach taken to tackle these issues centres mostly on the clarification and use of information-based ideas, and an investigation into the nature of the processes underlying perception. The primary objective is to demonstrate that information theory and extensions of it, and measurement theory are powerful tools in helping to find solutions to these problems.

The quantitative meaning of information is examined, from its origins in physical theories, through Shannon information theory, Gabor representations and codes towards semantic interpretations of the term. Also the application of information theory to the understanding of the developmental and functional properties of biological visual systems is discussed. This includes a review of the current state of knowledge of the architecture and function of the early visual pathways, particularly the retina, and a discussion of the possible coding functions of cortical neurons.

The nature of perception is discussed from a number of points of view: the types and function of explanation of perceptual systems and how these relate to the operation of the system; the role of the observer in describing perceptual functions in other systems or organisms; the status and role of objectivist and representational viewpoints in understanding vision; the philosophical basis of perception; the relationship between pattern recognition and perception, and the interpretation of perception in terms of a theory of measurement. These two threads of research, information theory and measurement theory are brought together in an overview and reinterpretation of the cortical role in mammalian vision.

Finally the application of some of the coding and recognition concepts to industrial inspection problems are described. The nature of the coding processes used are unusual in that coded images are used as the input for a simple neural network classifier, rather than a heuristic feature set. The relationship between the Karhunen-Loève transform and the singular value decomposition is clarified as background the coding technique used to code the images. This coding technique has also been used to code long sequences of moving images to investigate the possibilities of recognition of people on the basis of their gait or posture and this application is briefly described.

The greatest thing a human soul ever does is to see something, and to tell what it saw in a plain way ... To see clearly is poetry, prophecy and religion — all in one.

John Ruskin 1856.

'It's only a rough draft.'

List of Tables and Figures

Preface

The motivation for this project originally arose within the Control Group in the School of Electronic Engineering, Dublin City University[1], in the mid 1980s. At the time it was articulated as a desire to "close the control loop" in robotic manipulation using visual feedback. It was a natural extension of both the concept and application of the notion of *feedback* which is ubiquitous within control theory. What was envisaged was some kind of computer vision system capable of using three-dimensional information about its environment to guide a robot arm in assembly-type processes, and to do this in a relatively unstructured manufacturing situation. The justification for this approach was, and is clear [1]. It would allow greatly increased flexibility in the manufacturing process, with lower re-tooling, re-configuration and re-training costs. It would permit some component handling and feeding tasks, which could not hitherto be achieved by traditional means, either because of the complexity of the component or the small quantities required[2] [2,3]. It would allow a consistency of handling, or finish, and a level of accuracy and repeatability, to be obtained, which would be otherwise unachievable.[3] More recently, the convergence of Automated Visual Inspection and Robotic Manipulation [3, p.vii] and changing views on the role of inspection in manufacturing [4,5,6] have led to a greater recognition of the various tangible and intangible benefits of integrating vision in process control [7,8]. Put simply, visual input promised the potential for fast and unobtrusive gathering of information on a completely different scale to alternative means of sensing.

Even in the early stages there were some evident qualifications to the wisdom of, or need for, visual input in the types of environment considered. Custom assembly machines could for example be cheaper and/or faster in any particular case. Or re-

---

[1]Then the National Institute for Higher Education (NIHE), Dublin.

[2] This issue of small product quantities, and in particular batch modes of manufacturing has been identified by Batchelor [3, p.5] as one of growing importance to manufacturing industry because of increasing consumer discernment and changing tastes. Batchelor quotes an estimate, that even at the moment 75% of manufactured goods are processed in batches of 50 items or less.

[3]In some cases such consistency is much more important than any attendant labour savings.

engineering, perhaps coupled with new product technologies such as greater levels of integration, could eliminate some of the situations where visual servoing would be useful. But if anything, these issues have decreased in relevance in many manufacturing situations because of changing trends in consumer requirements. Nonetheless, what were then, and still are relevant, are the potentially huge computational overheads and unknown engineering, computational, scientific and even philosophical issues which would need to be tackled. But notwithstanding this, the likely benefits certainly dictated that these issues should be investigated and the research work leading to this dissertation began in earnest.

Initially much of this work concentrated on algorithmic issues, principally in the area of edge-detection and stereopsis[4], on the basis that these types of low-level operations are computationally very expensive and a necessary precursor to the less well-defined processes that would take place at higher, more abstract levels[5]. A strong influence on my thinking at this stage were David Marr's ideas about a computational theoretic top-down decomposition of the problem in terms of the different levels of description of vision as an information-processing task [9]. Much of my effort in this early stage of the project was dedicated to trying to come to terms with the implications of Marr's framework and to use it as a basis for our robot vision system. While progress was made on the identification of suitable efficient algorithms for carrying out these early processes on the input digitized video images [10], there still remained many outstanding issues. Principal among these were problems such as how outputs from different visual modules like stereo, motion, texture, colour, etc. should be integrated;

---

[4] At an early stage in the project a decision was made to use passive means for extending vision into the third-dimension, such as stereo or motion analysis, rather than the so-called active approaches involving projection of light patterns or using lasers. The reasons for this are outlined in a technical report [1]. Note the term "active" has since acquired a new connotation in computer vision in the sense of camera movement involved in effective exploration. It is used above in the sense of imposing structure on the lighting configuration. Curiously, rejecting the active methods in the earlier sense of "active", meaning structured lighting, leads to a situation where the extra computational overhead associated with less structured situations can be alleviated by spatially varying resolution (cf. foveation) and the combined ideas of fixation and camera movement of the term "active" in the more recent sense.

[5]This dichotomy between so-called low-level (or image) processing and more abstract processing or descriptions has recently been exploited as a means of extending the capabilities (higher-level task descriptions and process feedback) and flexibility (less constrained environments) of automated inspection systems. This is done by mapping the image-processing functions to dedicated hardware and software, closely integrated with the image sensor (the Image Inspection Ltd. *Intelligent Camera*), and mapping the more abstract levels into functions in a specially extended version of the *Prolog* language. See for example, Batchelor [3] or Whelan [8].

what learning strategies could be used so that the system could cope with the learning necessities of an unstructured environment; or how to avoid the problem of needing to put more and more information into the system as the environment was allowed to become less and less structured.

During this stage I was using some of the more accessible literature on neuroscience, psychophysics and cognitive psychology, mainly to cross-reference design parameters such as acuity, spatial and temporal bandwidths, sensitivities, coding criteria, etc. (See for example Levine [11]). However it soon became clear to me from this literature that there were major inconsistencies and flaws in applying the prevalent computer vision models to biological vision systems and vice versa. For example, the explanation of the properties of neurons in the early visual system in terms of edge and line detection were contrived at least. Many of the computational models of visual capacities such as shape from shading seemed to have no place either in the edge-detection-type early processing of computer vision models, or in the context of the properties of neurons being described within neuroscience. The explanation of biological vision systems in terms of the "semi-independent modules of perception" like stereo, motion, texture, etc. [9] was weak – there was little evidence from neuroscience for this particular division into modules, and what evidence there was for independent pathways supported a much more subtle division of labour. And the list could go on.

This could all be explained in terms of the stumblings of an immature science, groping for some certainties in the mass of either inconsistent results or irreconcilable experiments. Or perhaps the top-down design methodology proposed by Marr was simply too difficult to carry through – we had little idea of what we should be trying to achieve, let alone discover the constraints that would determine a unique solution as was advocated by him to be the proper way to proceed. But while I was unhappy with the situation where different scientific disciplines were coming up with seemingly quite incompatible explanations and models for what was effectively one phenomenon, I had no idea of where to look for alternatives. In fact, at the time I probably didn't even think to look for alternatives. As far as I was concerned our scientific tradition was a unified objective way of examining and discovering reality – one which, since the seminal insights of such pioneers as Bacon, Galileo, Descartes and Newton had allowed us to make unprecedented strides in understanding, subduing and controlling our environment, and that Marr's ideas were a thoughtful, broadly-based, and hopefully useful approach within this tradition. Our scientific tradition is, as Rosen says [12], a *canonical* account of scientific explanation: it rules that there is fundamentally only one way of looking at the world and it is that way — the fact that it involves several tacit hypotheses about the natural world is certainly not obvious.

In hindsight it seems that much of the trouble was caused not by lack of knowledge of the details, nor using the wrong models, but because the whole foundation to the experiments and the models, the whole philosophy of the approach, was at the very least limited, without its limitations being recognised. Now, philosophy and biology are not topics that usually appear in modern scientific or engineering literature[6] so why should a dissertation on computer vision be any different? Well it seems that in vision, and indeed more generally in AI, we have drifted out of the domain in which the assumptions of our scientific tradition are necessarily valid — our philosophical positions need to be carefully re-evaluated, and carefully looking at biology may help us to do this. By this I do not mean that we should try in any superficial sense to *copy* biological vision systems in the mistaken assumption that they are in some way optimal. Rather I am saying that by calling something artificial *vision* we are invoking a biological metaphor — one to help guide our implementations and our applications. In some cases, typical of machine vision, this metaphor may be inappropriate and biological details quite unsuitable for solving the corresponding problems. In other cases, where the environment is less constrained, the biological vision metaphor may be quite appropriate and useful in coming to terms with the issues that arise here — but it must not be used loosely. Attempts to force biological ideas into the representational framework of the domain of engineering and design typical of machine vision, ignore the fact that in relatively unconstrained environments this framework is no longer appropriate, and *a fortiori* is not applicable to biological systems. We must be clear about the requirements, restrictions, limitations and possibilities of the domain of design and engineering, and be clear about which of them are no longer valid when we cross over to discussing or invoking metaphors from living systems. This is

[6]Having been awarded my primary degree through a department of Pure and Applied Natural Philosophy I feel in a sense licensed, though not necessarily qualified, to discuss these issues: unfortunately they are usually not on the undergraduate curriculum of a modern physics or mathematics department.

required so that our engineering efforts are not hampered by attempting to achieve the impossible, nor that our biological explanations are contaminated by inappropriate analogies, when the assumptions on which they are based, are no longer suitable.

Needless to say, this all seems quite clear to me in retrospect, but at the time it was very difficult to articulate exactly what the problem was. However, one concrete manifestation which I suspected was a symptom of some sort of a problem, was the way the ideas of information and symbol were used in computer and biological vision, and in connectionism. The usage of the term information in particular, was loose, usually undefined, often inconsistent, and yet seemed to be quite evocative. So, much of my work in the middle stages of this project was involved in trying to figure out the status of the concepts of information and symbol in a range of related fields. Eventually, this search in turn led me to the position described in the previous paragraph.

Initial positive influences (as opposed to the simple dissatisfaction expressed above) which had the affect of my conversion to this view about the need to re-evaluate some very basic ideas, came from a number of sources. There was for example the very interesting multi-disciplinary work in biological vision carried out in the late 1980s by Livingstone, Hubel and their colleagues [13,14] which seemed to provide clear evidence for visual properties that could not be processed serially and hierarchically (as for example in Marr's raw-primal sketch, full-primal sketch, 2½-D sketch, 3-D representation hierarchy [9]), but were required to be processed in parallel pathways. Also in this and related literature there were growing revelations that information flow in biological visual systems was not all in an "upwards" direction towards "higher centres". In fact there were large scale projections from parts of the brain where the neurons' activity seemed to correlate with more abstract visual properties, back to the early visual cortex where the neural projections first arrived from the retina. Not only were these reafferent projections acknowledged to exist but there were substantially more of them than the afferent projections from the peripheral sense organs. They seemed to have been largely ignored simply because they didn't fit into prevailing ideas about what visual processing involved. A second influence was the careful analysis of the retina and early visual cortex described for example by Laughlin [15] and Barlow [16,17] which seemed to indicate that *information theory* in the sense of communication has much more to say about the function of the activity of these organs than the *information processing* of feature-detection. Another was the work of Linsker [18] which demonstrated that an information theoretic principle alone could cause a particular artificial neural system to develop very similar microstructure to that observed in the receptive fields of the early visual cortex. Related to this was the work on neural networks that showed that the so-called "end-stopping" neurons in mammalian visual cortex, which were explained as corner detectors in edge/feature-detection schemes, could in fact be better explained in terms of shape-from-shading[7].

Probably the most important influence on my thinking at this stage was the work of Michael Satosi Watanabe. A graduate of Tokyo University (Ph.D., 1933) and the Sorbonne (D.Sc., 1935) where he worked with Heisenberg, his background was in quantum mechanics, information theory, logic, and later philosophy. In his 1985 book [19] he sets out to look at the status of pattern recognition, both as a capability embodied in biological organisms and as a function designed into artificial systems, tracing its origins from the writings of the classical greeks scholars to modern neuroscience and computer science. This book provides the basic inspiration for the presentation here. In it I first discovered that there are alternative philosophical standpoints from which to view pattern recognition, and by extension, perception in general. I discovered that despite its resounding successes there are pitfalls implicit in the standard world view when it is applied in areas where its basic assumptions no longer hold, particularly so because these assumptions are not presented as assumptions as such, with alternatives, and with domains in which their validity might not hold. Rather they are presented as part of *the* way to look at the way the world *is*, with no relativities and no qualifications.

In addition to Watanabe, I found a number of other authors immensely valuable at this stage in pointing out what the important issues are and the way to possible solutions. I include here the extensive work of Robert Rosen [12,20,21] of which I have yet only come to terms with a small fraction; Roland Wilson and his colleagues whose work on symbols and uncertainty first led me to Watanabe; and the immensely lucid account of information by Fred Dretske [22].

---

[7]By definition, something is only a detector for a property if it responds only to the presence of that property and not to any other property.

There is some acknowledgement within pockets of the computer vision community that the problems described here are real and need to be tackled. Examples of the sorts of questions which have begun to be raised are:

What is the role of vision?

Does vision have a context?

Does vision have an input or an output?

Does it make sense to talk about a vision system?

What is the relationship with the application domain?

How can we classify different vision systems?

What is the relationship between autonomy and vision?

A decade ago such questions would not have merited a second thought. There is now a growing awareness that these issues are not as clearcut as they once seemed. Prompted by this change, a workshop within the ESPRIT BRA Working Group on vision entitled "Vision in Context" was held Killarney in September 1991 to begin to address the issues raised by these questions. A further workshop supported by the ESPRIT BRA Working Group entitled "Autopoiesis and Perception" was held in Dublin City University in August 1992 to examine a possible direction which would begin to supply answers. Complete answers are not yet forthcoming to most of these questions but at least now we seem to be asking the right questions.

It was in this context that I first became fully aware of the implications of a school of thought within which these questions not only made sense but were answerable, at least in principle. Much of what is discussed in this dissertation was stimulated by the framework and ideas put forward by Watanabe. Nevertheless, the work of Watanabe, important though it has been, was just a stepping stone for me to the more radical and more sophisticated *enactive*[8] viewpoint expounded in the extensive work of Humberto Maturana, Francisco Varela and their colleagues [23,24,25,26] over the last twenty or so years. For a number of reasons, the enactive viewpoint is not central to the development described in this dissertation, but it is used to evaluate the ideas presented, in the role of something like a perspective commentry. An alternative

---

[8]The work of Maturana and Varela is most closely associated with the term *autopoiesis* (literally meaning self-production). Notwithstanding this, autopoiesis is not actually the key idea of their work. Rather as Varela himself has said [26], the term autopoiesis has become "emblematic of a view of the relation between an organism and its medium". This view is a notion which he has recently tried to capture with the term *enactive*. A term used by Randall Whitaker to capture the same concepts is *autopoietic theory*.

approach to the issues discussed here, would be to use the enactive viewpoint as a guiding philosophy from the very beginning of the project. This is both possible, and quite likely to be fruitful, but such a radical approach was not necessary to point out some of the deficiencies or limitations of the conventional understanding of computer vision. A change which is incrementally different, might in fact be more likely to influence the conventional beliefs. Having said this, I am happy with the content of the presentation here for more personal reasons, because it reflects my intellectual development over the duration of the research. This is probably the better for having grappled with the difficulties of the problems before seeing the potential solutions. Certainly I am more in awe of the great intellects which have struggled with these problems and advanced our thinking on them so much.

Even though much of the details remain to be elaborated within the enactive approach, in it we finally find a single, solid, consistent and quite persuasive framework in which artificial and biological intelligence can be discussed and evaluated – not a framework in which the representational viewpoint is redundant, for there are domains in which it is relevant – but one in which its limitations are recognised, and in which more encompassing alternatives are available when required. Much work still needs to be done to complete the approach to understanding perception which in described here, but it is also time to begin a programme which treats vision purely within an enactive context. The fact that it has been possible to closely intertwine the two approaches in this dissertation is indicative of the strongly complementary nature of the two.

Because the work described in this dissertation is very much cross-disciplinary, it is felt appropriate to make explicit the nature and specifics of the contributions being made. The original contributions contained in this dissertation include:

- a clarification of the relationship between the domains of computer and biological vision;
- a review of the philosophical basis for work in computer vision;
- a proposal, supported by argument and evidence from the research literature that the fundamental unit or event of a perceptual system is a primitive observation or classification;

- arguments to support the view that Boolean logic is not a suitable algebraic framework for dealing with decisions about a relatively unconstrained environment and that non-distributive logics are more appropriate;
- support for the proposal that the development of autonomous systems with a perceptual component for operation in unconstrained environments should take place at an operational level of the systems' organisation and dynamics rather than at a representational or symbolic level;
- a clarification of the role of perception, and the status of the camera/eye and computer/brain analogies;
- a review of the research literature on the operation and function of the retina and particularly of the information-theoretic interpretation of this function;
- a review of some of the different interpretations of information, including ideas associated with Shannon and Gabor, and the notions of redundancy, entropy and structure;
- a clarification of the status of the terms "information" and "symbol" and their use in the context of biological and computer vision;
- a review of some of the more important theories of perception to date, particularly those which have had an influence on research and development in computer vision;
- a comparison between the causal framework of Rosen and the different types of explanation described by Varela and a description of how these might be used in the study of the dynamics of autonomous systems;
- pointing out the relationship between the Karhunen-Loève Transform (KLT), the modified KLT (related to Watanabe's "object-predicate inversion") and the Singular Value Decomposition (SVD);
- demonstrating and clarifying the coding possibilities of the SVD for data reduction, particularly in pattern recognition problems;
- the application of these ideas about the SVD, along with artificial neural network (ANN) classification, to the problem of inspecting solder joints on printed circuit boards;
- the application of the SVD and ANNs to the recognition of people through their motion when walking;
- an investigation of the interpretation of the activity of nodes in a simulated ANN in terms of symbol processing ideas;

- a review of the application of probability to pattern recognition, the nature of the inductive inferences involved and how the problem of inductive ambiguity can be overcome;

While specific contributions in terms of review of progress to date, clarification and extension of ideas, applications, and the advancement of new ideas, are contained in this dissertation, I think I will gain most satisfaction if, as a whole, it makes a contribution to the argument which clarifies what we really should be doing when we are carrying out research into vision, or building artificial vision systems. For this question, more than any other, is the question which has motivated me to do this work. Once I had seen for the first time, unaided, the effects of a random-dot stereogram, I dearly wanted to know what it really meant, how it came about, and how I could build something that would "see" it as I saw it. The solutions are not contained in this dissertation, but I am confident that at last, I am heading in the right direction.

List of Publications

The University's regulations requires that previously published material be referred to in the thesis, so for convenience a list of the publications which have arisen, either directly or indirectly out of this work have been gathered together here. They are as follows:

1. Byrne,N., **Murphy,N.** & McCorkell,C., "Comparison of Motion Estimation Algorithms", *Proc. of 7th Conf. of Irish Manufacturing Committee on Advances in Manufacturing Technology & Systems (IMC-7)*, Trinity College Dublin 1990, pp.190-201.

2. **Murphy,N.** & Gunning,J. "Automatic Feature-set Selection for Automated Visual Inspection", *Proceedings IMC-8*, University of Ulster at Jordanstown, September 1991, pp.671-685.

3. **Murphy,N.** & Byrne,N., "Long Sequence Analysis of Human Motion using Eigenvector Decomposition", presented at *ESPRIT BRA Workshop on Computer Vision for Surveillance Systems*, S. Margherita Ligure, May 1992, (paper in preparation).

4. Gunning,J. & **Murphy,N.**, "Neural Network based Classification using Automatically Selected Feature Sets", *Proceedings of the IEE Third International Conference on "Factory 2000"*, University of York, July 1992.

5. **Murphy,N.**, "The Causal and Symbolic Explanatory Duality as a Framework for Understanding Vision", in McMullin,B. & **Murphy,N.** (eds.), *Proceedings of ESPRIT BRA Workshop on Autopoiesis and Perception*, August 1992, Dublin City University, (submitted for publication).

Chapter 1

1 Introduction

1.1 Biology and Vision Engineering

There is no doubt that the aim of computer vision is to design and build machines that have at least some of the visual capabilities of people and animals to sense their environment – and this is not without justification. Little reflection is needed to see that it is a most astounding faculty, with a myriad of different aspects, and one which underpins so many of our human abilities. Furthermore, it is difficult to describe an application task in the engineering domain of computer vision without making use of the anthropomorphic terminology and analogies associated with seeing. Even the word *vision* itself immediately conjures up many of the various sensing nuances of our very personal interactions with our environment. Nevertheless, the relationship over the past three decades, between work in the areas of biological vision on one hand, and that of artificial vision systems on the other, has been quite chequered. In his posthumously published 1982 book [9], David Marr describes how initial excitement about the possibilities of computer vision generated by discoveries in psychophysics and neuroscience during the 1950s and 1960s was largely not followed through during the 1970s. The ideas proposed by Marr and his colleagues led to a renewed interest and cross-fertilization during the early and mid 1980s. Marr's own work was intended to be both a computational model for biological visual phenomena and, by definition, a potential description for artificial vision systems – he saw no distinction between these two different aims at the level of abstraction with which he was concerned. In practice, the relationship is far more detached. For example, current work on computer vision (see for instance [27]), uses biological results which largely date from this period during the early '80s, with few major advances happening in computer vision because of biological discoveries made in the intervening time.

Within the computer vision community, biological vision systems are frequently held up as an existence proof, sometimes even as an optimal vision system, on the basis that evolution has had millions of years to explore the space of possibilities. Yet despite this apparent esteem, usually only cursory reference is made to actual psychophysical capabilities or specific neural implementations in computer vision literature. Some authors, for example Levine, have argued for more positive links:

> *Everyone agrees that the problem of programming a computer to analyze and, what is more, to understand the content of pictures is extremely difficult. My view is that if we are expected to write algorithms to achieve these goals, it is incumbent upon us to know how humans and animals achieve this same function ... and to show the relevance of biological models to engineering systems. [11, p.xiv]*

Nevertheless, there are a number of reasons why positive interactions between the two fields have been more conspicuous by their absence over the years, than by their presence. There is the straightforward reason that few practitioners in either domain are technically competent in the other, and while some research groups have tried to tackle this by forming multi-disciplinary teams, it is a practice which is not widespread. Another reason, particularly for the lack of progress during the 1970s, is that, as Marr pointed out, the two fields have quite different aims. He claimed:

(i) that traditionally the emphasis in neurophysiology and psychophysics has been to *describe* behaviour (a phenomenological approach), rather than to *explain* that behaviour, so the results were of little use to the engineering and design domain of computer vision where the emphasis is on function [9, p.15], rather then mimicry;

(ii) that the approaches in computer vision which predominated during the 1970s were mostly either (a) "unashamedly empirical", with little analysis of performance or optimality, or (b) designed to be restricted in scope to "toy" problems with the hope of subsequent generalization: either way they were quite unsuitable for helping in the understanding of real biological vision systems.

What was needed Marr suggested, was a decomposition of the problem in terms of different levels of explanation:

- a *computational theoretic* level to make explicit "what is being computed and why" and to show that it is optimal;
- a level of *representation* and *algorithms* to describe how the computational theory could be implemented;
- and a hardware *implementation* level describing on what physical substrate the representation and algorithm would be realized.

In principle this attempt to distinguish different types or levels of explanation is a useful methodological discipline, particularly in organising *a posteriori* the details of a given model or system. In practice, the top-down strategy which Marr strongly advocated based on these levels, has had only limited success in providing solutions to computer vision problems. The reason for this is primarily because it is so difficult to think up a computational theory for many of the ordinary functions of vision. In fact we have only a limited conception of what the functions of vision really are at this stage. Equally Marr's approach has had only limited success in helping to understand observations within biological vision because having thought up a computational theory it is very difficult to relate this to the observed biological mechanisms at the realization level. For example, one of the things that seems to be missing, is not just a proper description at the computational theory level in Marr's sense, but rather a common framework, equally applicable to both biological and computer vision points of view, within which a computational theory might be formulated. The framework of set theory, Boolean logic and symbol systems implicit in Marr's particular computational theories of visual functions like edge-detection or stereo vision or spatial representations, is something which we argue here to untenable.

More decisive in the lack of success of the interactions between computer and biological vision than any of the reasons Marr gives, is something that he failed to realize (primarily because, as Marr says himself, his approach was an extension of the so-called representational theories of mind). It is in fact often the case that the engineering and design domain of computer vision, and the domain of biological vision systems' operation, so apparently closely related by the epithet "vision", really have very little to do with each other. More fundamentally, it is frequently the case that they actually involve contradictory, or at least incompatible foundations or premises. The fact that people *do* draw inferences about one domain and quite inappropriately use them in the other is unfortunate, though understandable. According to Varela, this is a situation which has arisen mostly because of a lack of clarity or precision in treating the relevant concepts. Having criticized the desire for *purely* operational descriptions (which are discussed below) as a remnant of the logical positivist view with its emphasis on methodological monism, Varela continues as follows:

At the other extreme, the vitalist attitude, and more importantly the computer-gestalt attitude, which take information as 'stuff', are equally misguided. The latter attitude is interesting for it has taken the same kind of methodological flavor implicit in operational descriptions, and applied it to a domain where it simply does not work. This is typical in computer science and systems engineering, where information and information processing are in the same category slot as matter and energy. This attitude has its roots in the fact that systems ideas and cybernetics grew in a technological atmosphere that acknowledged the insufficiency of the purely causalistic paradigm (who would think of handling a computer through the field equations of thousands of integrated circuits?), but had no awareness of the need to make ***explicit*** *the change in perspective taken by the inquiring community. [24, p.77]*

The use of the ideas covered by the general term "information" is something Varela singles out for particular criticism: an attempt to carefully develop an appropriate role for "information" is one of the central themes of this document. The change in perspective that Varela refers to, is the key to understanding the appropriate context for using the term "information", and the appropriate relationship between the domains of engineering and design on the one hand and biological systems on the other.

Arguably the only application area in which the general corpus of computer vision has achieved real success to date, in terms of effective and valuable utilization, is machine vision. This is the application of image processing techniques to automated inspection and to a lesser extent to robotic manipulation. The concepts, applications, techniques and methodology of machine vision are as far removed from biological vision as any other technical domain such as control or signal processing is. Only the vaguest of similarity remains: they both use a combination of light, the reflectance properties of surfaces and suitable optics to achieve their respective functions. The issues and concerns of machine vision are only addressed somewhat obliquely in this dissertation, in the sense that one of the aims is to clarify the relationship between biological vision and computer vision and to determine under what conditions it is appropriate to draw conclusions about one on the basis of the other. For the reason given above, this shows that computer vision in certain contexts[1] is more closely related to biological vision than it is to the context usually referred to as machine vision. Machine vision, in

[1]There is no name like "machine vision" for these contexts yet, though the working term of *artificial perception* is sometimes used here.

particular, is primarily concerned with achieving certain results in a particular contextual situation where the mechanics, lighting, optics, sensing and processing can all be *configured* to optimize what can be achieved. Even though it is often not made explicit, the archetypical application of a general computer vision system is as a subsystem of a system which is effectively an autonomous system, and not just that, but which is capable of coping successfully with an environment which is effectively out own. But the level of specification and control typical of machine vision, is anathema to the functioning of an autonomous biological system. Despite the *apparent* sensing similarities or analogies, despite the "vision" common to both, the two domains are quite incommensurate. On the other hand, some general conclusions about say the logical calculus intrinsic to sensing and measurement systems might apply equally to machine vision systems as well as both biological and artificial perception though this is not discussed directly. The primary concern here is in dealing with the situations in computer vision where understanding biological vision should be of use (in other words dealing with relatively unconstrained environments) and examining biological vision to this end.

Virtually every technical domain borrows terms from another context, by way of analogy, or resemblance, and gives them meanings peculiar to the context of that technical domain. Words such as "group" (in mathematics), "work" (in mechanics), "information" (in communication theory), or "chaos" (in system theory) spring to mind as terms having a technical meaning related to, but somewhat different from their everyday connotation. One problem with computer vision, and possibly AI generally, is that often the words are actually *meant* in the technical domain in the sense of the term in the original domain and even when they are not, the distinction is not made clear[2]: there is seldom a precise definition of what is intended by the term. In a general sense this is what Varela is driving at in the quotation above. The conceptual distance between machine vision and biological vision already mentioned means that this is not normally a problem for practitioners of machine vision. The wide variety of processing and analysis tools which have been devised bear little relationship to the processes evoked to explain biological functioning. That is, the tools of machine vision are useful, precisely because they are like spanners to fit particular nuts: once we can

[2]Examples of this are terms like "edge", "shading", "texture", "recognition", "motion", "shape", "feature", or even the term "information" itself.

clearly identify the type of nut in question, there are a range of different spanners available to deal with it, depending on the circumstances[3]. The problems seem to become qualitatively different however once we allow the vision system's environment to be substantially less constrained. Now, developing a relationship between artificial vision and biological vision becomes potentially more relevant and the pitfalls of confusing incompatible ideas from the respective domains more real. The argument is not that all artificial "vision" systems can benefit from interaction with ideas from biology: this is likely to be the case only where the problem domains are in some sense commensurate. They are likely to be for both computer and biological vision. They are not for biological vision and machine vision.

Unfortunately, while the problem domains of computer vision, involving as they do, far less constrained environments do indeed make biological vision systems more relevant, they also bring us further away from the domain of engineering and design congruous for example with machine vision. In particular they bring us further away from a domain in which representations are an appropriate abstraction for dealing with the domain and further away from a domain in which, in a sense to be explained below, information is fixed.

1.2 Objectives

The five human senses have been described as conduits for the perception of our environment. More specifically and more correctly, they are a means for regulating our behaviour in our environment. Of the range of sensory modalities, sight is probably the most striking in its capacity and utility. Visual proficiency has enormous value for survival: it provides the ability to remotely and quickly detect structures and events in the surroundings in a most direct and unobtrusive manner. But to talk about *the* visual capacity is misleading. Firstly, it is one which is represented in the animal kingdom in such an incredible variety of forms and levels of sophistication that there can clearly be no *necessary* visual sense, no particular physiological structure or

[3]This is intended to caricature rather than trivialize the situation. In fact many more processing and analysis tools have yet to be devised and the really useful developments may be in how the tools are selected, combined and alternated in a context dependent way.

functional capacity[4] which *must* be present in order to identify a capacity as visual. Each species has uniquely adapted visual sensitivity and its implementing mechanisms to their particular requirements in their ecological niche. Secondly, what we normally refer to as *the* sense of vision, in humans say, seems to actually consist of a number of sensory capacities which happen to share some common mechanisms (e.g. the eye) because light is one link in each of the sensory-motor loops involved[5]. These are two important points. They are the first steps towards understanding that, contrary to common sense, there is no objective reality full of information which our eyes (meaning our brains) simply pick up, and other creatures to a lesser extent depending on their ability. In a certain sense we *construct* our own reality — our own world — not in isolation to our environment[6] nor logically consequent on it, but instead logically compatible with it and with our own organisation. As Cariani says in a colourful use of metaphor[28, p.xvi], it simply doesn't make sense to say that a seagull has a more *real*istic model of the world than a lobster because the former's interactions are more sophisticated. They simply face radically different challenges. (See also [24, p.68]).

Notwithstanding the diversity of functions that can be described as perceptual, if there are *common elements* underlying the processes of perception or its developmental mechanisms, then knowledge of such elements would be invaluable to the overall understanding of biological vision and its artificial implementation. The aim of the research culminating in this dissertation has been to discover what, if any, these common processes are and how they can be used.

---

[4]Other than the trivially necessary function of being sensitive to light within a particular range of wavelengths.

[5]To illustrate this point consider the "sense" of sound perception in the bat. It is actually a medium in the two separate sensory-motor loops of spatial perception using sonar principles and sound generation for something like mating calls.

[6]In the context of living systems Varela [26] makes a clear distinction between the notions of *environment* and *world*. That is, he distinguishes between the environment *of* the living system as it appears to an observer, which he calls simply the *environment*, and the environment *for* the system which is defined in the sense that the system distinguishes itself from what is not itself, and that only exists in that mutual definition, which he calls the system's *world*. This is a key point because it draws the distinction between what is important or relevant for the observer who has access to both the system and its environment, and what is important or relevant to the system in its world. This of course only becomes germane if it is realised that the observer's world and the system's world are not identical, and that in fact by definition there can be no single objective world commonly apparent to both.

It was in an effort to tackle these issues that the questions of what perception is; what it means to perceive – to be aware of the world around us; what the mechanisms of perception are and how they develop, were posed as a point of departure. It is not possible to properly deal with these without confronting some of the classical debates which have occupied the minds of philosophers since the time of Plato and Aristotle. In fact we find that aspects of certain philosophical positions are valuable, particularly in pointing to alternatives to the standard received scientific mindset. So also we find that a careful examination of aspects of biological operation, of information theory and pattern recognition, and notions of the relation of autonomous systems to their environment, provide clues which begin to throw light on these issues. These are the primary fields of investigation with which we are concerned in this dissertation.

1.2.1 Understanding the nature of perceptual processes

One of the objectives of this dissertation is to propose that there is a fundamental level of description which is the basis of, and common to, all perceptual or sensory processes – regardless of the sensory modality, the species of the organism, or whether it is a natural or artificial system – and to support this assertion on the basis of a broad spectrum of evidence and arguments in fields ranging from philosophy, psychology and neuroscience to physics, mathematics and computer science. More particularly, a primary objective of this dissertation is to show that information theory and its extensions, and measurement theory are very powerful tools which allow us to go a long way towards the aim of discovering what processes underlie visual perception. What is presented here is not a complete theory or model. Rather it could be considered as an essential primitive or framework which should be incorporated or reflected in full theories or proper models. This is meant in the sense that Boolean logic is reflected in the design and operation of a digital computer but the operation of any particular program is not reducible to Boolean logic – the programmer and the system designer have an essential input. More specifically it involves an attempt to clarify the role and operation of symbols and the relationships between what are often characterised as the domain of semantics and the domain of syntax [28].

It is proposed here that there is a level of description of systems in the context of which it is appropriate to claim:

(i) that the fundamental operational unit or event of a perceptual system is itself a primitive type of *observation, measurement*, or *classification* event;

(ii) that perception involves many of these primitive 'observations' arranged as an interacting system or network – not as a single layer of transducers or detectors, but where the output of one "signal-to-symbol" transition can be used as the input of another;

(iii) that these 'observations' may not necessarily be implemented as discrete physical or physiological units, but may result from the dynamic functioning of an underlying structure at a different level of description;

(iv) that without an aspect of its operation which can be described in terms of these 'observations', it does not make sense to ascribe perceptual functionality to a system;

(v) that any particular percept involves the simultaneous activation of many many primitive observations spread across the system or network. There is nothing else required to 'look at' any output of the perceptual mechanism for perception to occur. The act of making sensory information explicit is perception;

(vi) that the primitive observations involve inductive generalization – a classification process not unlike that in classical pattern recognition – which overcomes the ambiguity inherent in implicit information;

(vii) The classification process is one of measuring a predicate – applying a universal property or general concept to the input stimulus data for, *or relative to*, each primitive observation. Each particular measurement output is one and only one of a mutually exclusive set of possible outputs, making an explicit decision that the input stimulus data satisfies a certain general property or concept.

Exactly what the basis is for these claims and the details of what is meant by them is teased out in the following chapters.

Like most modern theories of perception, the ideas described here were originally conceived in terms of the assumption that visual data derived from scenes in the external world implicitly contain information about the physical structures and events contained in the scene and some of their properties. Perception then, is the process of making this information explicit. The development for the author of the realisation that there are problems with this way of looking at the world was a very slow process. In hindsight, and particularly in the light of the epistemological position espoused by Maturana and Varela [23], it is possible to see clearly why the description of the function of visual perception in terms of "picking-up" and processing information from the environment, and the role played by information in this context is inadequate. Most of the elements of this shift of position were already in position when the author became aware of the work on autopoiesis, but the ideas of Maturana and Varela represented a very welcome opportunity for support, clarification and development.

1.2.2 Understanding the nature and role of information

There are at least four distinct views on how the notion of information should be interpreted. The conventional approach to interpreting biological vision systems and experimenting with computer vision systems has been overwhelmingly dominated by a representational view of information [9]. Even more recent connectionist approaches, though embodying a substantial change in viewpoint, have only involved a change of the *type* of representation to one of a distributed nature [29]. The notion of information as being constructed and co-dependent rather than instructional and referential is an interpretation based on the more embracing viewpoint of the complementary causal descriptions and symbolic descriptions playing clearly defined interrelated and dual roles rather than mutually exclusive or even muddled roles [24]. A third, and very different view of the notion of information is that captured by the Shannon and Gabor theories of information. It has been argued that these are not actually about information at all, being concerned with something more akin to the what we think of as the properties of a signal [22]. Finally there is the everyday semantic connotation of the term information which has motivated the previous three uses, though it is not captured by any one of them (nor necessarily should it be expected to). It is possible to extend the theory of information (in the Shannon sense) to accommodate semantic aspects of the everyday meaning of information and we discuss this in chapter 7.

These different interpretations have not arisen in a void. They are part of the history of the development of the endeavour to understand the human mind, which is what we turn to in the next section. Teasing out what is intended by these four different contexts and interpretations for information was considered both as an objective of this dissertation and as a methodology to support the primary aim of understanding perception. So, from the original aim of understanding, at a very fundamental level, what the essential nature of a visual capacity is – what vision is – the specific objectives of the research described in this dissertation have arisen. In other words, understanding vision means coming to understand first the nature of the process of perception itself, and the nature and role of information, in what is often referred to as an information processing system or task.

1.3 Cognitive Science

The development of computer vision as seen by Marr is described in section 1.1 above. With the benefit of developments over the decade since Marr's ideas were originally published, it is now possible to get a clearer view of how these developments fit into the broader development of cognitive science. The scientific effort to come to terms with the nature of intelligence, perception, thought and behaviour is usually labelled Cognitive Science. According to Varela [30] there have been four major stages in its development over the last forty years: *cybernetics*, *cognitivism*, *connectionism* and *enaction*. This latter term is likely to be unfamiliar to many readers. It is a term which has been recently coined by Varela to describe the radical departure in the understanding of biology and cognition for which he, along with Humberto Maturana, has been primarily responsible.

Cybernetics

The original programme in the new science of cognition, which spanned the decade from 1943 to 1953, was called cybernetics. It was a wide-ranging cross-disciplinary effort to create a complete science of the mind. It achieved in its time many far-reaching results, including the application of mathematical logic to the study of the brain, the invention of computers, substantial contributions to systems theory, control theory and information theory, and the demonstration of the possibilities for self-organising systems. It came to an end largely because the principal participants had died, or their views and interests had diverged.

Computationalism

The second phase in the development of cognitive science, began in 1956 and is still the dominant position in the field. It goes under many general titles, including cognitivism, computationalism, artificial intelligence (AI or GOFAI[7]), and so on. It also has many specialized sub-areas like expert systems, robotics, computer vision and speech recognition, focussing in particular types of problem domain or modalities. The methodology is generally of a *top-down* nature and used in both analysis (cognitive psychology, computational neuroscience) and synthesis (artificial intelligence). The central ethos of this approach is that cognition is *defined* as rule-based manipulation (computation) on symbolic representations, where the meaning of each symbol is made to correspond to an external item in a restricted well-defined domain. With this definition, what was originally a tentative idea or metaphor within cybernetics, is elevated to the status of a full-blown principle. Information in this context is considered as an objective quantity associated with objects and properties in the world. It can be detected, processed, and used to build internal representations of the way the world *is* external to the organism or system [9].

Connectionism

While the connectionist or emergence approach also has its origins in the early work on cybernetics, for various well-documented reasons it has only recently developed a level of adherence sufficient to allow it to challenge and complement the dominant cognitivist position. The methodology is usually *bottom-up* and is characterised by distributed processing using simple sub-symbolic components and by self-organization leading to global network coherence. The self-organisation is usually realised in terms of adaptive connections (between nodes) which, affected by "experience," change the strength of these connections according to certain rules (eg. the Hebb rule and its variants, or error back-propagation). In terms of synthesis its successes have been

[7]Acronym for Good Old-Fashioned Artificial Intelligence

primarily with lower level cognitive capabilities which cause most difficulties for the cognitive approach, such as recognition, association, and memory. It may be possible to integrate the cognitivist and connectionist positions by embedding symbolic levels of description in an underlying distributed system though only limited effort seems to have been put into this problem so far. (See the discussion in section 8.5 below).

Much of the emphasis and success within the connectionist community to date has been on the distributed and bottom-up aspect of the connectionist approach associated with the so-called PDP[8] models [31,32]. The basic epistemological position is still representational, though the form and construction of the representations is quite different from the cognitivist approach [29, p.252]. In this case it is a global state or performance of the system which is related to meaning in some chosen domain rather, than the value of a localized symbol. Nevertheless, there is still an observer external to both the system and its sphere of operation, and this observer provides the connection between performance and meaning. That is, there is always a teacher to supervise the learning phase of the network model and the model comes to reflect more or less accurately and successfully some of the cognitive concepts of the teacher. Even the measurement of accuracy and success are dependent in the final analysis on the teacher.

1.3.1 Representational view of reality

The objectivist position implicit in both the GOFAI and PDP traditions centres around the commonsense idea that the world as we experience it is independent of the knower. It does not distinguish between my world and the environment of someone or some living thing like me. The problem of perception, within this approach, is to find algorithms or mechanisms which will allow this absolute reality to be captured from the flux of generally ambiguous visual information that is available. Knowing, within this approach, is the act of "duplicating" what is already there outside the knower, using the senses to convey information to construct the appropriate representations. What is represented is a correspondence between symbolic units in one structure (the representation) and symbolic units in another structure (our world or frame description).

---

[8]PDP is an acronym for "parallel distributed processing", but in fact stands for a more restricted programme than the general label suggests.

Rosen [12] claims that the problem of the representational approach is inherent in a view of the world that dates back to Newton's attempts to develop mathematical models of the world. He illustrates what the problem is in the following way. Consider

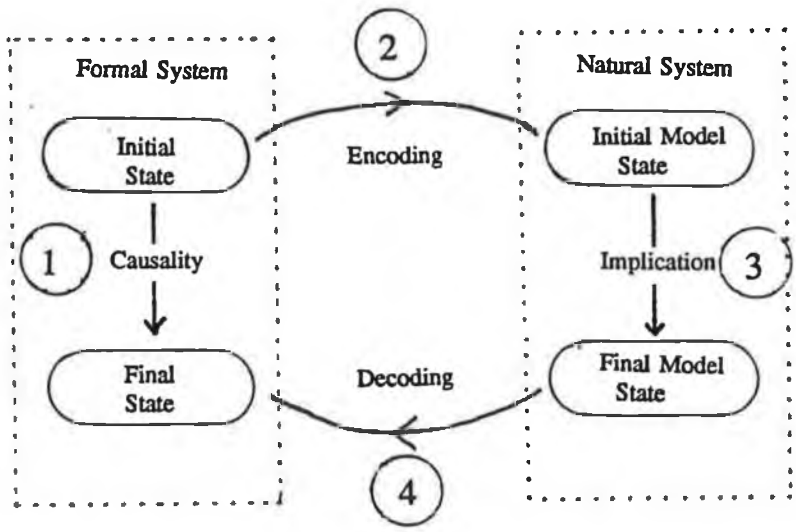


Figure 1. The Newtonian modelling scheme. Adapted from Rosen [12].

the modelling scheme shown in Figure 1. On the left hand side is represented (some part of) the natural world, for which the supposition is that events are related by definite laws or relations – by some causal order. On the right is represented some formal system with elements which are governed by mathematical relations based on logic and implication. Relations are established between these two "worlds" by encodings and decodings, where objects or events in the natural world are (arbitrarily) identified with elements of the formal system, and vice versa. With this modelling relationship one should always obtain the same result by observing the causal order unfold in the natural system (arrow 1) or making predictions via the encoding, mathematical inference and decoding arrows 2,3 and 4 (given that one uses a suitable mathematical model). There are a number of implications of this type of modelling scheme, which Rosen draws out within the causality framework discussed below. But the primary problem with this view, according to Rosen, is that the encoding and decoding arrows that Newton originally *posited* with the ideas that culminated in the so-called Newtonian world view, have become axiomatic, and therefore *invisible.* Thus the natural world is considered as something which in principle, can be modelled to an arbitrary accuracy with a Newtonian-type formalism (dynamics) in which each element in the natural world plays a fixed causal role and can be represented by a fixed symbol in the formal system – a system in which *information* is thus fixed.

Hence the reference to "symbolic units in another structure" above. Once this is clarified, it is possible to see that:

(i) the encoding arrows are established by us and thus arbitrary; they are not a necessary consequence of the nature of the world itself;

(ii) the encoding arrows are not something that can be "picked up" from the natural world side of the diagram in the sense that we might expect an intelligent machine to "learn" about its environment; and

(iii) that systems other than Newtonian-type formal systems (dynamics) could be represented on the right hand side.

The ideas dealt with by Rosen involve encodings of the natural world and formal models of this world. Nonetheless they are useful because they exhibit clearly, ideas which suffer from the same shortcomings as the representational view of reality, perception of that reality and our internal conception of that reality. Varela approached the same problem from a different angle and brings the idea further:

> *Both in cognitivism (by its very basis) and in present day connectionism (by the way it is practised), it is still the case that the criteria for cognition is a successful representation of an external world which is pre-given, usually as a problem solving situation. However, our knowledge of activity in everyday life reveals that this view of cognition is too incomplete. Precisely the greatest ability of all living cognition is, within broad limits, to* ***pose*** *the relevant issues to be addressed at each moment of our life. They are not pre-given, but* ***enacted*** *or* ***brought forth*** *from a background, and what counts as relevant is what our common sense sanctions as such, always in a contextual way. [30]*

The problem with the representational approach is that there is no way, within the system supposed to construct these representations, of ever obtaining the appropriate assignment of correspondence – there is no independent access to the supposed external reality for any given living system. The cause of this problem according to Varela is a confusion of different levels of explanation: it is the confusing of notions proper to the domain of an observer (or strictly an observer community) whose vantage includes both the system and its interactions with its environment on the one hand, with notions proper to the operation of the system on the other. These are different phenomenal levels, and links, if any, between them can only be established by someone external to both the system and its environment.

1.3.2 Self-organisation – emergence

More recently however the connectionist approach within cognitive science has been instrumental in forcing a re-interpretation of the role of individual neurons in the analysis of biological neural systems. This reinterpretation involves a move away from the information processing and representation role exemplified by, for example, the theories of hierarchial visual processing of Hubel and Wiesel [33], Barlow's "grandmother cell" [34], or anything like Marr's hierarchial structure of primal sketch, 2½-D sketch and 3-D representations. In its new role the neuron is seen as belonging to large transient ensembles of coherently active neurons where no single neuron is responsible for, or even restricted to, a single aspect of perceptual experience [35,36,37]. What is more, the reafferent neural projections from higher cortical areas to the early sensory cortex (which far outnumber the afferent projections from the sensory organs to the sensory cortex), after being mostly ignored in theories of cortical information processing heretofore (largely because they didn't fit into the direction of information flow implied by representationalism) are now being recognised for the role they play in the emergence of global cortical phenomena. This alternative approach to connectionism is very different in its philosophy from either the GOFAI or PDP concepts. It emphasises the *self*-organising properties of connectionist models rather than their representational possibilities [38,39,40,41]. Adaptation – if it can be called that – takes place without the benefit of supervision: the activity of the systems is determined by the structure of the system itself. These developments are already pointing in the direction of Varela's fourth stage within cognitive science, the *enactive* viewpoint, although the enactive perspective is generally concerned with broader issues than the properties of particular models and the relationship to their inputs from an environment. It seeks in fact to revise the very roots of our epistemological stance, not eliminating the notion of representation but making clear its restriction to well-defined situations, described *a priori* by an external agent or observer.

The enactive perspective

The single most important assumption of the dominant cognitive science tradition is that the world as we experience it is independent of the knower [30] and, consequent on this, that the primary task of the perceptual part of a cognitive system is to capture an accurate representation of aspects of this world. Even though this tradition found inspiration from the classical Newtonian view of physics, and as discussed above, shares some of its limitations, it has remained despite the radical overthrow of the Newtonian world view and its conception of ontological reality. Thus, even though the tenets of relativity and quantum mechanics mean that there are fundamental deficiencies or flaws in the Newtonian picture, these are mostly only manifest in cases of extreme speeds or relatively small sizes, and are mostly perceived as not affecting the "common sense" ideas of the world embodied in Newtonianism. Unfortunately, the very objectivity which is essential to the Newtonian picture, which is responsible for its success, which becomes untenable at very small (or quantum) sizes, and which is the basis for the representational view of the world, is also the single most important feature which does not carry over to the representational explanation of perception. Whatever the merits of objectivity in the analytical sciences, it is increasingly being seen as incompatible with reasoned views on the nature of perception and the relationship of an organism to its environment.

The opposite extreme to the view that the nervous system objectively maps *the* external world, is the notion of solipsism – that what is perceived depends solely on the structure of the organism itself. The fact that cognitive phenomena cannot be understood in terms of a world that "informs" us because there is no mechanism that makes this informing process possible, makes the non-objective extreme no less unpalatable [25]. Fortunately there is an acceptable view, intermediate between these two extremes, which has been articulated in the extensive work of Maturana and Varela. Basically this is that knower and known arise in a process of mutual specification. Neither the structure of the world nor the operation of the observer are pre-given – they are co-determined by a history of cognitive interaction, neither logically preceding the other, but still logically compatible. There are two main issues implicit in this stance: the type of system that can participate in this co-determination of a constructed "reality", and the methodology of explanation which maintains distinct, the type of description appropriate to each phenomenal domain. We examine these in turn.

1.4 The Control/Autonomy Duality

The drawing of a distinction between a system and its environment is the most fundamental act of system theory [24, p.84]. Where we, as observers, put the emphasis of this distinction largely depends on our perspective or our purpose in making the distinction. If we focus on the internal operation and organisation of the system we are putting its environment into the background and relegating interactions with the environment to the status of perturbations. We are also emphasising that the properties of the system arise from within its own structure (the interactions of its components) with the environmental perturbations possibly *triggering*, but not *specifying*, the ongoing operation of the machine. On the other hand, if we focus on the environment, the system is treated as a simple system with given properties and its interactions with its environment constitute a part of its definition. The natural problem type arising from this latter view is the control of the behaviour of the system by utilising the constraints with its environment. This latter case is essentially the subject matter of control theory. The former case where the system is emphasised is the domain of autonomous systems theory. Varela uses the terms autonomy and allonomy to distinguish between these two different ideas.

Allonomy, literally meaning external law, implies the regulation or control of a system from outside [24, p.xi]. Interactions between the system and its environment are "instructive" and constitute part of the systems organisation. Unsatisfactory results from these interactions are errors. The organisational paradigm is usually formulated in terms of *input-process-output* and is organisationally open. This view of a system is suitable for the domain of design where an observer specifies by its use what the environment should be and how the system ought to use or participate in it. In other words it involves a representational viewpoint with the observer or designer specifying the appropriate semantic correspondences between (formal) descriptions of the environment and the (formal) organization of the system. *Autonomy* on the other hand literally means self-law, implying the internal determination and regulation of the system's operation. Interactions with the system are seen as perturbations which are non-instructive, and independent of the definition of the systems organisation. Varela uses the metaphor of conversation to describe our interactions with autonomous systems[9]. Unsatisfactory results from these interactions are represented by mis-understanding. The organisational paradigm is one of *circularity* and the system is organisationally closed. Information is considered as constructed and co-dependent where the outcome of perturbative inputs and outputs reflects structure attributed both to the environment and to the internal operation of the system arising over a history of continued operation of the system (and hence viability in its environment). In particular, information is not something which is simply picked up from the environment.

In addition to these complementary ways of making the fundamental distinction involved in systems theory, it is important to distinguish between the organisation of a system and its structure or realisation. The precise definition of a machine cannot be in terms of a list of its parts, or its potential use or purpose -- rather it must be by these, plus a description of the permitted inter-relations of the machine's components. A machine's *organisation* is the set of "relations that define the machine as a unity, and determines the dynamics of interactions and transformations it may undergo as a unity" [24, p.9]. There is no connection with materiality in the definition of a machine's organisation: it does not specify properties of components that allow the realisation of a machine as a particular concrete system. This closely parallels the idea of *relational biology* described in the 1950's by Rashevsky (see eg. [12])[10].

[9]The metaphor of conversation is an antidote to the sort of "double-think" which it is easy to fall into in considering the development of "intelligent" or autonomous systems. The question arises as to how these systems might be controlled. Could we suspend their intelligence or autonomy while we program them with instructions or are they only going to be partially intelligent or partially autonomous, always subject to some master designer, engineer or programmer? To what extent is this latter case compatible with autonomy?

[10]According to Rashevsky, described in Rosen [12, p.172] "we are interested in the organisational features common to all living systems; and in their material structures only in so far as they support or manifest these features. Therefore we have heretofore approached organisms in precisely the wrong way; we have abstracted out, or thrown away, all those global organizational features in which we are really interested, leaving ourselves with a pure material system that we have studied by purely material methods, hoping ultimately to recapture the organization from our material studies... Why do we not , in effect, *abstract away the physics and the chemistry*, leaving us with a pure organisation, which we can formalize and study in completely general abstract terms; and recapture the *physics* later through a process of *realization*." It is important to make this distinction between organisation and realisation because the physics (including the molecular biology) involved in the realisation of real organisms is logically compatible with the organisation of the organism or biological system but *not logically prior* to it and therefore quite useless in defining the organisation.

On the other hand, a machine's *structure* is the set of actual relations that hold between the components that realise a particular instance of a machine in a given space, and is determined by the properties of these components. Finally the *use* to which a machine is put is not a feature of the organisation or even directly the structure of the machine, but rather the domain in which the machine operates. That is, it belongs to our description of the machine in a context wider than the machine itself – the domain of observation (or design). This clarification leads us directly to the next topic.

1.5 Descriptions and Explanations

In his 1979 book "Principles of Biological Autonomy" Varela sets out to lay bare the relationships between "a systems *identity*, its performance in its *interactions* with what it is not, and how we *relate* to these two distinct domains" (p.xii). Already, embedded in this statement of the issues of concern is a pervading circularity which is the cause of much of the confusion of levels implied in objectivism. This is the case, for implicit in our act of description of a system and its environment are the peculiarities and particularities of the nature of the relationship between ourselves and *our* environment. More explicitly:

> *... the study of autonomy and [a] system's descriptions in general cannot be distinguished from a study of the describer's properties ... the system and observer appear as an inseparable duo. [p.63]*

By expressing an interest in the nature of perception – often inappropriately considered as generating a description of one's environment – we are immediately embroiling ourselves in these issues[11].

In spite of the circularity we have to start somewhere. Here we will start with the notion of description, but in the particular role of explanation – our explanations within a scientific community. In this context Varela draws a distinction between symbolic (or communicative) explanations and operational (or causal) explanations. The difference lies in both their form and use. Operational explanations are assumed to be defined in terms proper to the domain in which the systems that generate the

[11]Varela tackles these issues head on by attempting to answer the question: "How do we come to have items such as, say, frogs or people, of whom we can say that they perceive other things?". See for example [24, section 16.5.2].

phenomena in question operate. The purposes of operational explanations are those of prediction and manipulation. Symbolic explanations are assumed "to belong to a more encompassing context in which the observer provides links and nexuses not supposed to operate in the domain in which the system that generate the phenomena operate" [p.66]. In this case the purpose is quite different. This type of explanation is for communicating an understanding between members of a scientific community. The fundamental bases of operational explanations are nomic or law-like relationships. The fundamental basis of symbolic explanations is order or pattern, and it is the observer who establishes the connection. But it is not meant by this that the causes or laws, often so-called "laws of nature", are in some sense superior by being more remote from the observer, more objective. Both types of explanation are

> *modes of description adopted by enquiring communities for some intentional purpose ... and they specify modes of agreement and thus coupling with the environment. [p.77]*

The basic argument of autopoiesis is that all biological phenomena can in principle be reduced to a particular type of network of nomic relationships in some material domain. In this operational description notions of purpose, message, information or code play no *causal* role. But this is not the whole story: it may not be desirable or practical or useful to reduce every aspect of biological phenomena to operational descriptions. It may be very useful for *our* purposes to abstract or parenthesize a number of steps in a causal chain, choosing to ignore the operational connections in favour of more convenient descriptions. This is what Varela claims is at the base of all symbolic descriptions: a process of abstraction rooted in the emergence of certain "coherent patterns of behaviour" to which we *choose* to pay attention.

> *The fact is that information does not exist independent of a context or organisation that generates a cognitive domain, from which an **observer community** can describe certain elements as informational and symbolic. Information, **sensu strictu**, does not exist. (Nor, of course do the 'laws' of nature). [p.78]*

Thus, using information in a causal or operational role, e.g. relating behavioural regularities (in the domain of interaction between a system and its environment) to structural change (within the system), is a confusion of levels. The behavioral regularities are only available to us as external observers with simultaneous access to the operation of the system and its interactions with its environment. They reflect our operations and they are not operational for the system. The system does not have independent access to the nature of the structure of the environment, nor does it necessarily share our categorization of the environment.

So, what *is* a valid symbolic explanation? Well according to Varela symbols in natural systems are characterised by two main features: internal determination and composition. Internal determination refers to the claim that an object or event can be considered as playing the role of a symbol

> *... only if it is a token for an abbreviated nomic chain that occurs* ***within the bounds of the system's operational closure*** *... whenever the system's closure determines certain regularities in the face of internal or external interactions or perturbations, such regularities can be abbreviated as a symbol, usually the initial or terminal element in the nomic chain [p.80].*

In addition, symbols which are syntactically composible to yield valid combinations seen to confer selective value on the organism to which they belong, though we would caution that in the light of ideas discussed in chapter 6 below, the syntax involved might not necessarily be based on Boolean algebra.

All in all, the distinctions between different types and purposes of explanations emphasized by Varela, and the careful epistemological balance portrayed by the notion of enaction, both provide a boundary condition, within which to contain our developing understanding of the concept of information. In different ways they warn us off imbuing information with an excessive and undeserved semantic concreteness characteristic of both representationalism and objectivism: what Varela calls the "temptation of *certainty*" [25, p.16].

1.6 Summary of Chapters of Dissertation

The subject of this dissertation is sensory perception. Visual perception is taken as the paradigmatic case and even though little attention is directed here to issues or examples drawn from other sensory modalities, there is no reason in principle why the ideas addressed here might not also be applied to them. An assumption implicit in taking visual perception as a subject of research, is that there is at least some sense in which visual perception exists and is instantiated. In other words, there is an implication that there is, or could be, some part or subsystem of both biological organisms and presumably artificial systems, which could be largely described as participating in perception-type behaviour, processes or functioning. It is quite possible that this idea of a perceptual (sub-)system is only a first approximation, or that the idea is undefined without taking into consideration other non-perceptual aspects of the functioning of the organism or system. Nevertheless, in both of these cases, it should be possible to make progress in the explication of the nature of such perceptual (sub-)systems. This is the stance taken here. If this turns out to be an inaccurate approximation, or an invalid definition, then one would expect that the progress in the explication of perception should encounter insurmountable paradoxes or difficulties of one sort or another. That there remain difficulties at the conclusion of this dissertation is beyond doubt. But there is no indication yet that these are insurmountable, so at this stage anyway, the assumption of a perceptual (sub-)system is still a useful one.

Having assumed that it makes sense to talk about perceptual systems, it must make sense to discuss the nature, properties or precesses that uniquely allow them to be identified as perceptual, i.e., give them their perceptual function. We described above the assertion that at one level of description, the processes underlying any type of sensory perception can be understood in terms of a theory of measurement involving networks of primitive "observation" events acting on signals. These primitive observations receive their input signals from other parts of the system or network which are almost exclusively "local" relative to the site of the primitive observation, so there is no global "teacher" or "monitor" "looking at" the results of each primitive observation and determining if it is correct or useful. That is, there is no independent access to reality, relating the possibilities or outcome of an observation to a particular environmental context. Nor is there some central site of cognition, some higher centre, where the results of all the individual observations are brought together to form a single integrated concept.

Linsker [18] draws an evocative picture of this type of scenario, which helps to illustrate the sort of relationships involved, both in the immediate processing from moment to moment, and in the long term adaptation or development which leads to the particular nature of the processing occurring at a particular point and time in the system. Consider a person at a certain level of management in an organization, whose job it is to make the most informative summary possible of the data received by them each week. The particular type of data received depends on the external environment in which the organization exists and operates, the structure of the organization, (including what level this person is on and what other levels and functions exist within the organization), and many other constraints. Over time, this person comes to realize that a certain way of representing information (Linsker gives the example of graphical plots involving various variables) is most useful for summarizing the data for the management level immediately above. If this person interacts with other people within their particular layer, they may either (a) try to avoid unnecessary duplication by giving almost orthogonal summaries, if everybody does their job diligently, or (b) be forced to provide several almost independent copies of the same summary, if several people are lackadaisical about their work. In any case, the presumption is that the conscientious workers will try to ensure that their layer is as informative as possible. It is likely that eventually a set of work practices and job procedures will gradually come into operation, so that this person's layer will end up carrying out a processing function, without the workers in this layer needing to know either the goals of the entire organization, or what information is considered most important by the more senior managers in the layer(s) above. Furthermore, there is no need for any higher layer to try to reconstruct the raw data from the summary – rather the higher layers simply need the ability to discriminate the relative value of different actions. If the required information has been lost in one of the intermediate layers this cannot be done. Equally, if workers in the intermediate layers are not aware of the remote high-level goals they cannot be expected to know exactly what information might be safely discarded, and so it might reasonably be expected that they try to preserve as much information as is practical with the given constraints.

Linsker introduces these ideas in the context of a principle of maximum information preservation within a layered perceptual system. The corporate metaphor is not perfect but serves well to illustrate the key points of the local nature of signals arriving at a point, and the purely local primitive processing functions which are isolated from any direct contact with an external reality, either by way of direct contact with the source of the raw data or with the high-level goals of the system.

The particular processing model of concern to Linsker is that of a multi-layer feed-forward network with linear weighted summations at nodes and Hebb-type adaptation of the weights. There are no non-linear thresholding or classification operations at his nodes. The example above illustrates the local nature of the signals arriving at the processing points (the worker's desk or the network's nodes), but says little about the nature of the processing at each of these points. A claim made here is that for the purposes of describing the system's *operations* it suffices to consider the processing that is applied to the local convergence of signals. For the purposes of explaining the system's operations in a wider context that includes its environment, it may be useful to relate this processing to the original raw data[12], or to the high level goals of the system. But this is something that we do as observers, and it is not operational for the system. One of the basic claims of this thesis is that the fundamental processing unit or primitive in a perceptual system is a type of primitive observation or classification on the basis of a *local* confluence of signals only. These decisions, or classifications, or primitive recognition events can also be related to the environmental context of the system, (by us as detached observers) and in particular can be related to the original raw data, giving rise to the idea of a general concept associated with each primitive observation or decision. In other words, the values of these local signals are the basis for a decision that the general concept corresponding to each primitive observation has a certain value – an output value for the primitive observation is selected, or prepared, or decided upon. This value is itself a signal and may provide the input for other subsequent decisions or observations. This value is the output state of a local decision or classification[13], but we may relate it to a certain general concept, or classification, or measurement of a property, in the larger context which includes the system's environment. The general concepts which the measurement or observation processes apply, by operating on the incoming data, are the key to understanding the need for primitive observational processes. This notion of general concepts already has a long history which we can use to throw light of the properties of the type of perceptual systems under consideration here. We discuss the notion of general concepts, or

[12]The raw data case is not unlike the idea of receptive field used in studying biological vision, which is explained in section 3.5 below.

[13]It might seem, on initial reading that the juxtaposition of signals converging and diverging within information processing networks, with classification events based on the local convergence of signals would imply some sort of generalized perceptron-type neural network model is being proposed here. In fact, it seems likely that even though it is possible to describe both the transport of signals within the sensory network, and the primitive recognition events of sensory perception, in information theoretic terms, these are in fact descriptions which involve two different levels of explanation. These ideas are discussed further in the concluding chapter.

universals, as interpreted by philosophers through the ages in chapter 2 below. We find that we can come to an understanding of the notion which is useful both in mechanical pattern recognition and as a basis for a theory of perception. Because the process central to a primitive observation is a process of classification, or recognition, we examine pattern recognition from the point of view of what it involves and how it becomes possible. The process of pattern recognition requires the overcoming of the inductive ambiguity inherent in grouping and generalization, so we discuss the status of inductive inference. These three subjects: the problem of universals, pattern recognition and inductive inference applied to the notion of concepts form the core of chapter two.

The concepts embodied in the measurement process of each primitive observation are not innate although the architecture and gross connectivity of the hierarchy of observations is likely to be genetically specified. The actual concepts used in a mature sensory system seem to have arisen from developmental processes which operate somewhere on the continuum between (i) resulting from information processing principles inherent in the architecture (particularly its pattern of high local connectivity and ordered maps between different regions of the hierarchy) and (ii) resulting from adaptation to correlations in data from the external world. Chapter 3 begins to explore exactly what is meant by this in real biological perceptual systems. This fairly extensive chapter is important because it embodies the belief that looking at biological visual systems can be useful for understanding the nature of perception in general, even artificial implementations if that is possible. It does nonetheless attempt to achieve this aim in the cautious manner counselled in the preface. The retina (or strictly the retinas of many different species) is the main focus for this chapter because its structure and operation are relatively well understood and thus allow us to meaningfully compare different views of its function. Much of the early work on the vertebrate retina was carried out on the cat and interpretations based on these results, which historically have had an important influence on computer vision, are examined and compared with some more recent work on the primate retina. The methodology of examining the constraints which seem to be responsible for the present forms of the retina, is seen as a powerful tool giving clues as to why particular structures are relevant and what their function might be.

In chapter four we begin to tackle the issue of exactly what is meant by information, and what the relationship between the different meanings are. The emphasis is mostly on quantitative notions of information, but this groundwork is a necessary precursor to a cautious consideration of semantic aspects of information. The quantitative notion of information is traced from its roots in physics in the form of the concept of *entropy*, through its development into Shannon's information theory of discrete symbols on communication channels, Gabor's "minimum uncertainty" representation of analogue signals and Watanabe's structure function for pattern recognition. We also describe attempts to give quantitative descriptions of aspects of human vision using information theoretic ideas. Possibly the most valuable knowledge of a biological vision system would be an understanding of its development, because this would make explicit what actual goals the system is trying to achieve and the criteria it is trying to optimize in the search for these goals. Initial, but exciting work in this area is described in chapter 4. The basic conclusion is that aspects of the structure and processing of the early parts of biological vision systems can be understood in terms of attempts to *maximize the flow of information* (or minimize the equivocation) through the system. Finally, in this chapter we begin the discussion about the relationship between the quantitative and semantic aspects of information.

Historically, theories of biological visual perception have had some influence on ideas within computer vision. The major approaches suggested to date to explain neural function in the visual system are reviewed in chapter five. The notion of an invariant is closely linked with the idea of a symbol[14] and since the endpoint of a process of visual perception has traditionally been assumed to be a symbolic representation of aspects of the outside world, much effort has been put into the identification and computation of invariants. The "ecological" approach to visual psychology puts a strong emphasis on the idea of invariants. Gibson claimed that even higher-order invariants are directly detected, by the visual system "resonating" to them. These issues are the subject of section 5.3. There then follows an extensive discussion on Gabor filtering and Gabor codes, the relationship between the Gabor representation and redundancy, and finally possible explanations for the properties of cortical neurons and

[14]The direct relationship between symbols and the concept of invariance derives from the fact that because a symbol set is a generalisation within the set of possible signals, a single symbol will represent many symbols. In particular, signals can undergo certain transformations and still generate a certain given symbol. The symbol is said to be invariant to these transformations.

aspects of cortical processing in terms of Gabor coding. Again information theoretic notions play a very valuable role.

Measurement theory, classification theory, pattern recognition and discrimination theory are all based on the same fundamental idea: that of assigning labels to processes in such a way that processes that bear the same label are considered to be *alike*, and processes bearing different labels are considered to be different. We have already begun this discussion in chapter 2 and we try to look at it more carefully and in more depth in chapter 6. We review the different aspects of pattern recognition, and the notion that experience is vital to the ability to generalize which is intrinsic to pattern recognition. The various sources of ambiguity inherent in the generalization process of any act of pattern recognition are discussed in a probabilistic context. Shannon's major innovation in his 1948 paper [42] was to introduce the geometric representation of communication signals in terms of multi-dimensional spaces. We examine here how geometric representations can be particularly useful for understanding what is involved in pattern recognition. The really interesting notion is that in different situations, different *types* of geometric representations may be appropriate, and these in turn have different underlying algebraic structure. In particular, subspace representations of sampled data which code contrast (or relative values) rather than absolute values of intensity do not have the usual Boolean algebraic structure associated with commonsense views of the world in terms of objects and their properties. Propensity theory and fuzzy theory are related ways of dealing with ambiguity about objects or their properties, and propensity theory in particular seems to give insights into the nature of perception. This is discussed, along with the idea of evaluating predicates in a "test" or measurement and the properties of this process.

In chapter 7 the two related strands of measurement theory and information theory are finally brought together and used to discuss the possible processing structure and function of the early visual pathway, particularly the cortex. Before doing this, the notion of a "signal-to-symbol" transition is discussed in some detail, and problems with the usual interpretation of this concept are pointed out. Another view on this idea of a transition between two different representations with different properties, involving some sort of a generalization or classification or measurement, is provided by Dretske, and we examine his development of a semantic theory of information for this reason. The principal conclusion of this chapter, and the dissertation as a whole, is that the information theory/measurement theory interpretation of perception is much stronger than the relatively arbitrary computational theoretic explanations, for a number of reasons. It helps to avoid the pitfalls caused by hidden assumptions about our own values or caused by the prejudices of our own observables. It is a quantitative theory which eliminates the need to postulate arbitrary representations and processes. It is nearer to the actual operational level of an organism than symbolic descriptions and it can be used to understand most aspects of perception in any system, human, animal or artificial without requiring the invention of a new computational theory in each case.

In chapter 8 we develop some of the ideas introduced in chapter 6 and describe their use in an industrial inspection problem. The principal development here is the demonstration of a relationship between the Karhunen-Loève transform and the singular value decomposition. The details of the coding and recognition phases of the system are described and results are presented for both of these.

Chapter 2

2 The Philosophy of Perception

2.1 Introduction

Somewhat surprisingly (at least to the author) the work of the classical Greek philosophers Plato and Aristotle, written almost 2500 years ago, still has valuable things to say about the matters of concern here. We do find that it is necessary to be somewhat selective about the notions that are useful. Even the context in which they are used is often far removed from the speculations of the original authors. Nevertheless the underlying concepts are frequently sufficiently powerful that they can still clarify many of the issues that confront us. We describe here how ideas from both Plato and Aristotle can be used in this way. Another reason for the importance of these and later related sources is that without necessarily realising it, much of our present view of the world and the way we deal with it, particularly in western culture and even more particularly in science and engineering, is strongly influenced by them. Specifically, Aristotle's ideas about matter, objects and their properties are strongly reflected in commonsense views about the world and in classical (Newtonian) physics. In its detail, Newtonian mechanics was an overthrow of the earlier Aristotelian ideas about mechanics, but the essence of the Aristotelian view of the world still remained[1]. This discussion brings us to the foundations of modern pattern recognition where we see the impossibility of achieving anything without either explicit or implicit heuristics which in some way capture our values and our way of seeing things. The importance of this notion of value or usefulness is discussed. The subject of inferences, deductive and inductive, is one which is often misunderstood but has an important bearing on the issues of hypotheses and truth and reality. It is examined here from a probabilistic point of view. As a prelude to the later discussion on the nature and function of perception we deal with the more obvious fallacies that have caused trouble in the past and discuss the extent of what can be assumed about perception. But more generally, one of the major issues which has motivated philosophical debate since the time of

[1]In chapter 8 we show how an alternative point of view on objects and their properties can be very useful in certain coding or pattern recognition applications, though this application is not central to the development of the ideas in the remainder of this thesis.

Plato and Aristotle until comparatively recently is the controversy of the status of universals. Ideas from this controversy are at the core of the notions being proposed in this dissertation and it is to this that we turn first.

2.2 The Controversy of Universals

As a point of departure for the discussion that follows, let us consider first the naive position that the world can be thought of as consisting entirely of distinct (in a material sense) well-defined *objects*. Objects in the world can be said to share features (or predicates like colour, shape and so on) with other objects. It is in the nature of most such features that they can characterize indefinitely many objects, and because of this, these features are often called *universals*, where particular object is an instantiation of one or more universals, usually very many. So, in this sense, a universal is a concept or a general idea. For example, if we make a claim such as "this is a wheel", then the pronoun "this" designates some particular object while the noun "wheel" designates the general property of "wheel-ness" or the general concept of being a wheel, in other words a universal. Similarly an object with the property of being a wheel might also be designated by the adjective red, meaning that in addition it has the property of "red-ness". The principal problem is to describe the status of the universals and the particular objects [43,44]. The standpoint taken on this issue has important consequences for pattern recognition in particular, and as we shall see, for perception in general.

In the classical philosophical problem of the relationship between universals and particulars there were two principal positions (though many shades of opinion within each). In what must be an unlikely use of terminology, *realism* is the view that ideas, Forms[2] or universals are the only true reality, belonging to a world beyond matter and appearance. The world of appearance has only a temporary, illusory existence. Furthermore, the human mind can only apprehend the particular by virtue of its being able to apprehend universals – the notion of universals prior to the objects. (This latter notion is crucial to the development below. It provides the key to the understanding

[2]The modern meaning of the word 'form' in this context is that of shape, structure or figure, etc. In Greek philosophy the Form played the role of the epicentric concept of an ideal representation or definition of a class or general idea. In an Aristotelian adaptation of Plato's philosophy, "the 'Form' or 'Idea' is an ideal, real object that exists in an eternal world different from the actual world of daily experience. Individual concrete objects of this latter world are nothing but defective copies of the 'Form'" [19, p.7].

of perception which is presented here). On the other hand, one of the common denominators of *anti-realist* views, including *conceptualism and nominalism*[3] is that the human mind *can* directly apprehend the particular [19, p.52].

2.2.1 Realism

Plato is generally acknowledged to have been the founder of the realist view. While he and Aristotle differ considerably over the understanding of the term "Form", they both agree that the Form in the sense of universal is some*thing* real. The standard interpretation of Plato's work is that he treated universals as real things (Forms or ideas) separate from their instances[4] (particular objects) and independent of human understanding. That is, a particular object does not have a *real* existence, only a deceptive, temporary, illusory existence – what really is, is the (supposed innate) Form or idea, which is a state, a function or a meaning. The particular object is only an imperfect copy of the ideal Form.

From this point of view, the sensory world of experience has no reality, but the eternal world of form has reality. A particular object belongs to a class corresponding to a universal because it "partakes" in the archetype or form corresponding to that universal. Plato himself, however, encountered great difficulty with exactly what this notion entailed, passing through at least three different phases involving different, though related uses of the term Form [19, p.47]. It is possible that much of the subsequent criticism of Plato's Forms arises because of the later Aristotelian bias that forced the idea into the role of substance, which Plato did not intend [p.93]. In fact Plato introduced the notion of Form mainly for the domains of abstract concepts (such as goodness, kindness, bravery, and so on) and of mathematical concepts (such as numbers and geometric shapes), so that application to concrete objects (as in wheel or car) may be an over-extension of his notions:

> *The Form is not a perfect object in the best of worlds but rather the essential nature or functional meaning of the objects covered by the same name. [p.47]*

---

[3]Popper uses the term *essentialism* as a name for any (classical) position which is opposed to nominalism, especially for the theories of Plato and Aristotle [44, p.20].

[4]Aristotle is also credited with holding the realist viewpoint although he denied that universals are objects or separate from their instances, instead claiming that they are real things which exist just *by* being instantiated – the notion of universals in the objects.

In this sense then, the interpretation of Form as a universal is not quite appropriate, and ironically, may owe more to an Aristotelian interpretation of Plato than to Plato himself. Not surprisingly in view of the idea of an eternal world of Forms, in modern parlance the term *Platonist* is usually associated with the reality or truth of abstractions (particularly mathematical ones such as numbers, sets, or propositions etc[5]).

Let us set aside though for the moment the status in terms of reality, truth, or origin, of the Forms or ideas playing the role of universal in Plato's theory. The really crucial notion as far as understanding perception is concerned, is the claim that the human mind can only apprehend the particular, by virtue of it being able to apprehend universals[6]. Regardless of the specific interpretation of what Plato intended, this debate is nonetheless useful because of the emphasis it places on general concepts, properties, or universals at the expense of concrete objects, and particularly because of the above claim. One of the aims of this dissertation is to examine this claim in the enactive context outlined by Varela [30] and to use it as a starting point in the development of a theory of perception [45].

To the modern western mind, immersed as it is in a culture which shares much with Aristotle's philosophies, the assertion that the world of experience is unreal and the world of ideas is real, is completely antithetical to "normal" modes of thought. The following example may help to clarify the notion of reality intended:

> *Think of a geometrical figure like a 'triangle'. A triangle drawn on paper is not a real triangle as defined by geometry. Because we can speak of deviation from, or approximation to, a real triangle, we cannot deny the reality (in a certain sense) of a geometrically defined triangle, which is the*

---

[5]The term Platonist in mathematics is usually meant in the sense that what a mathematician does from the Platonist viewpoint is to discover these abstractions, rather than view them simply as the formal consequences of a particular formal system.

[6]By emphasising this, our intention is not to claim that the mind can *directly* "apprehend" a universal like "dog-hood" or "wheel-hood". Before getting to predicates as abstract as these many levels of less abstract primitive observations (or measurements of predicates, or evaluation of properties) would typically be activated in a normal perceptual experience. But at each stage the basic operation is envisaged as being the same – one of using locally available signals about less abstract decisions in the perceptual network as the input for a decision based on these. Even though the words "network" and "system" are used here, it is not intended that the primitive observations or classifications described should be thought of in the sense of the nodes of an artificial neural network because it is likely that the level of description in terms of primitive observations is a level above the level of physical realization (which was referred to by the terms structure and organization in section 1.5).

Form of a triangle. An actual triangle drawn by human hand is a poor imitation and not a true triangle. [p.47]

A similar shift from substance to function is manifest in modern quantum physics. Here, experimental results have forced the exclusion of the notion of self-identity and substance (in other words concreteness) for elementary "particles". Properties of elementary particles are no longer considered as fixed attributes independent of physical observation, but as variable quantities which are dependant on, and affected by observation. Instead of saying: *particle P is in quantum state Q;* the description must be rephrased as: *quantum state Q is occupied by a certain number of the particle of a certain kind.* The roles of object and predicate are interchanged [p.94]. This analogy between measurement/observation in quantum mechanics and measurement/observation in perception is elaborated in Chapter 7 below. The idea of object-predicate interchange is useful both as an algorithmic device in certain types of pattern recognition problems (see chapter 8), and for forcing a rethink of the nature of observation in perception.

One further point of note in Plato's writing which is related to pattern recognition, is his use of the term "paradigm".[7] This changes during different periods and three distinct uses have been noted, consonant with the different uses of the term Form mentioned above. These are the use of the term paradigm to mean

(i) a general concept in the sense of Form as used above;

(ii) a general concept in the sense of a perfect model to be imitated, and

(iii) a particular example of a concept [19, p.47].

It is interesting in this context to compare the various modern meanings of the term "pattern".[8] The usage of the term paradigm strongly parallels the modern *epicentric concept* of pattern recognition. It also motivates the important notion of paradigmatic symbol in the extension of pattern recognition to general perception.

---

[7]*Paradigm* is used in this section in the sense of an individual object, or exemplar, standing for a class, paralleling Plato's use of the term. This is distinct from the Kuhnian sense of an ideological theory or approach to scientific problems for which the German term 'Zeitgeist' is probably more appropriate. In fact Watanabe describes Kuhn himself as regretting the widespread use of 'paradigm' in an ideological theoretic sense as "a rather unfortunate fad" [19, p.9]. Nevertheless this (latter) sense of the term paradigm is deeply ingrained and well understood in the scientific community and so can be quite useful. Which meaning is intended here will usually be apparent from the context.

[8]Usages of the term 'pattern' include (i) design, (ii) model, template, plan, (iii) regular way of acting, (iv) person or thing worthy of imitation, (v) a sample (includes not only the one imitated but those which imitate); any object qua a sample of a class; a typical case, exemplar, archetype.

2.2.2 Aristotelian duality

Aristotle, also a realist (meaning that he accepted the reality of universals), recognised the reality of particular objects, which he argued, consisted of two real elements – *Matter* and *Form*. Matter endows real existence to an object, and represents "potentiality" in the sense that a plank of wood has the potential to be worked into a chair. Form determines the essential nature or functional meaning of an object. It represents "actuality" in the sense that regardless of the matter involved, say wood or metal, Form is what determines whether the object is a chair or a table. Aristotle then, believed in two types of substances. The basic reality is the primary substance – a concrete object provided with attributes. The secondary and less important substance is the universal to which the objects belong, labelled by the same attributes. He thus denied that universals are objects[9] or that universals are separate from their instances. This leaves us with the fundamental assumption of Aristotelian philosophy which is based on his obsession with *substance*: what exists is (an enumerable number of) particular objects with fixed attributes or predicates. This should be contrasted with the epistemological position of Plato, that what exists is the Form of which we have an innate idea [19, p.443].

For modern pattern recognition and computer vision, for the 19th century version of physics which underlies popular intuitive ideas of science, and for much of the so-called "western" mode of thought, this view of particular objects and object-predicate duality is deeply ingrained. It pervades everyday language in the form of the "subject plus predicate" pattern of thinking. It finds its most concrete modes of expression in Boolean logic and in the probability calculus of Kolmogorov (see [46, p.341]). However, just as almost a century before, the exclusive validity of Euclidean geometry was challenged, so in the first thirty years of this century a number of challenges arose to the truth of the law of bivalence, or Boolean logic[10]. These two pillars of western thought had stood for over twenty centuries.

---

[9]As discussed above, Plato himself may not have claimed otherwise, but certainly Aristotelian interpretations of Plato's ideas have.

[10]See the relationship of Boolean logic to the object/predicate idea, in the discussion on the Frege principle and the propensity logic in Chapter 6

Louis de Broglie was the first to point out the incompatibility of the probability calculus based on Borel sets with quantum mechanics [19, p.518, 47]. In 1932 John Von Neumann [48,49] pointed to language with its framework of ordinary (Boolean) logic as the problem in understanding quantum mechanics.[11] He proposed the construction of a "logical calculus" in contrast to the concepts of ordinary logic as the solution to the problem of interpretation [50, p.271ff]. It seems that this quantum (non-distributive and anti-Aristotelian) logic may also be an appropriate framework for understanding perception. This concept of non-Boolean logic and Watanabe's notion of object-predicate inversion, display a strong affinity to the Platonic ideas discussed above, and are an explicit rejection of the common sense notions of object and predicates which are strongly Aristotelian in character.

So, for our purposes, the most interesting aspect of these Aristotelian ideas is that they represent a classical and commonsense position which deeply permeates current thought and practice in cognitive science, particularly AI. Recognising their origin, and their manifold manifestations (in for example, Boolean logic, probability theory and classical physics) allows us to begin to see the possibilities of alternatives to these modes of thought (or mathematical models), and the likely implications of alternatives.

2.2.3 Aristotelian causality and complex systems[12]

One further aspect of Aristotle's ideas is also quite useful in showing very clearly within systems theory, the implications of moving outside the usual Aristotelian (and Newtonian) framework. In answer to the question of why an object or artifact is the way it is, Aristotle attributed four different and inequivalent causes — four different ways of saying "because". That is, if we consider an object as the "effect", then its *material cause* is the matter comprising the physical manifestation of the object; its *formal cause* is the shape (form), plan or blueprint for the object; its *efficient cause* is the act of construction or the processes which shaped the object to its present form; its final cause is the reason for, the goal fulfilled by, or the use of, the object. However, in addition to their classical usage these causal ideas can be used as a useful

---

[11]See the discussion on logos and mythos in Zukav [50].

[12]The discussion in this section, while a direct development on the Aristotelian ideas met so far, is somewhat tangential to the primary discussion on universals and realism, and may be read separately.

framework for understanding, not only objects but also systems. Consider, for example, the following definitions [51]: the *material cause*[13] is the passive receptacle on which the remaining causes act; the *formal cause* is the essence, idea or quality of the thing concerned; the *efficient cause* is the external compulsion that bodies have to obey; the *final cause*, for a machine, is its use, aim or purpose. With these more general definitions, it is possible to relate this causal framework to the operational/symbolic distinction made by Varela which is discussed in section 1.5 above. The material, formal and efficient causes belong to the operational description of a system – they all involve categories or relations within the phenomenology of the operation of the system. The final cause on the other hand, which can be interpreted in terms of purpose or use, does not pertain to the machine's operation – it is not a feature of its organisation. Rather it belongs to our description of the machine in a context wider than the machine itself – in other words, Varela's symbolic explanation.

Rosen [12] is even more explicit about these relationships and uses this causal framework to directly interpret the dynamics of systems. Consider for example the dynamical system description in terms of the rate equations:

$$\frac{dz}{dt} = \psi_g(z, \beta(t))$$

where $z(t)$ is a state (or phenotype) vector
g is a system or species (genome) vector
$\beta(t)$ is a vector of environments (inputs, forcings or controls).

If the "effect" is the state $z(t)$ of the system at a given time t (cf. the notion of phenotype), the *material cause* is the initial state of the parameters in the state space $z(t_0)$; the *formal cause* is the type, form or identity of the system labelled by coordinates in a function space; the *efficient cause* is the operator that transforms the initial state to the current state, and which depends on the organisation of the system and its environmental inputs, i.e. the operator:

[13]It is important to distinguish *materiality* (involving the properties of components that define them as physical entities) and *material cause* as defined, which has very little to do with matter.

$$\int_{t_0}^{t} \psi_g(\ldots,\beta(\tau)d\tau$$

The notion of a *final cause* plays no role in this Newtonian-type formulation of a system's dynamics[14]. Rosen's claim is that the Newtonian paradigm only applies to those systems for which the categories of causation can be segregated into mathematically independent structures, and there is no category of final causation within this paradigm, as in the example above. This class of systems is referred to by Rosen as *simple systems*[15] and not all systems can be reduced to this Newtonian form.

Consider, for example the rate equations for a general dynamic system where the environmental controls and genomic labelling are temporarily omitted

$$\frac{dx_i}{dt} = f_i(x_1,\ldots,x_n)$$

Now consider the quantities

If u_{ij} is positive, then an increase in x_j implies an increase in the *rate of production* of x_i. That is, x_j is an *activator* of x_i. Similarly, if u_{ij} is negative, x_j can be described as

---

[14]We can, for example, only talk about *control* if we have some external observer setting a reference point for the system to attempt to attain. Such a reference or set point cannot be something intrinsic to the organisation (in this case the dynamics) of the system itself. If it were, the system would be useless for control purposes as it would always try to reach some internally determined set point, regardless of external input.

[15]Rosen uses the term *simple* here in a very particular sense: he means that it is possible to define a particular set of system observables (a state space) and a set of differential equations over these observables (a velocity vector field on the state or phase space), which if they initially satisfy some criterion for approximating the dynamics of the system, will continue to do so for an arbitrary length of time. However complicated the model is, it is essentially a finite and fixed description of the system. In information terms, once the empirical process of determining what the appropriate observables and real or abstract "forces" are, no further information is required by an observer, to correct or maintain the approximation offered by this model. Rosen claims that in general, this is not true of even the least complicated living or biological systems and thus refers to these as *complex* systems. In other words, while it may be possible to construct a multitude of partial models of the Newtonian type for biological systems which can approximate the behaviour of the systems, these approximations are only local and temporary. As the "complex" system develops in time, the discrepancy between what the "complex" system is actually doing, and the predictions of the "simple" Newtonian model, grows in an unbounded fashion. When the discrepancy becomes intolerably large, it becomes necessary to replace the initial "simple" model with another, typically involving both different observables (which are not necessarily functions of the original ones), and a different set of dynamics. An observer trying to predict the system's behaviour needs to regularly get more information about the actual system behaviour in order to maintain the accuracy of his short term predictions. It is in this sense that it can be said that information is *fixed* for a Rosen-simple system, just as Varela argues it is for a case where the representational paradigm is appropriate.

$$u_{ij}(x_1,\ldots,x_n) = \frac{\partial}{\partial x_j}\left(\frac{dx_i}{dt}\right)$$

an *inhibitor* of x_i.

Now, there are many situations where this type of activation-inhibition description is more appropriate than a rate-equation description, typical examples being biochemical, neural and ecological systems. Here, it is often not possible to determine the absolute value of certain variables and the "forces" acting on them, only the differential excitatory or inhibitory inter-relationships between variables. In fact, given cartain types of activation-inhibition or network descriptions, a corresponding dynamical system description may not even exit. If we have a description of a system in terms of u_{ij}'s we can only go to a rate-equation formulation if the differential form for each i

$$\omega_i = \sum_j u_{ij} dx_j$$

is an exact differential. For $n>2$ this differential form is exact only if

$$u_{ijk} \equiv \frac{\partial}{\partial x_k}(u_{ij}) = \frac{\partial}{\partial x_j}(u_{ik})$$

If u_{ijk} is positive, then x_k enhances or potentiates the effect of x_j on x_i and we can call x_k an *agonist* of x_j. Similarly if u_{ijk} is negative we can call x_k an *antagonist* of x_j. For arbitrary systems there is no special reason why the condition for exactness of the differential form should hold. When a differential form cannot be integrated to give an equation involving the x_i's only, it is referred to as a *non-holonomic* constraint. (There are other types of non-holonomic constraint, which do not involve differentials, but instead limit the motion of the system to particular parts of the state space. See for example [52]). Each such equation of constraint between the x_i's and the dx_i's can be used to eliminate one degree of freedom of velocity but not the corresponding configurational coordinate in the phase space. Because we cannot obtain a rate-equation formulation for systems involving such non-holonomic constraints, we cannot describe the system in terms of separate categories of causation as in the so-called Rosen-*simple* Newtonian-type formulation described above. Components of the system may play more than one causal role at a given time, as is typical of systems with a circular organisation[16], systems which Rosen explicitly describes as *complex*. In particular there is no such thing in this case as a set of states which are assignable to the system for once and for all[17]. Also, these non-holonomic constraints are examples of the type of regularities that an external observer might describe as symbolic in Varela's terms. Perhaps this type of situation is characteristic of systems which display non-trivial metadynamical organisation [53].

The point of this section is not simply to add to the development of the idea of universals or the realist/anti-realist arguments. Nor does this section attempt a unification of the pseudo-Platonist universal used in perception, and Varela's views on autopoiesis and perception, with Rosen's ideas on systems. The objective is to illustrate the fact that the philosophical arguments of epistemology and ontology have close parallels in more concrete and more familiar fields. It may be that such a unification is possible, or that at least the different results arise from similar sources, but I suspect that the mathematical tools required to demonstrate this are not yet available.

2.2.4 Anti-realism

Returning to the discussion on universals, if *realism* is the view that universals have real existence, then the diametrically opposed view is referred to as *nominalism*. This is the notion that the universal is a name (or word) without any real existence. *Conceptualism*, holding the middle ground between these two extremes, is the view that the universal does not exist in the real world, but has a real existence as an idea in our mind. One way of viewing the realist stance is that the mind can only apprehend particulars by virtue of their ability to apprehend universals. The common denominator among the spectrum of anti-realist views which oppose this position is

---

[16]Note that by circular organisation we do not simply intend systems with feedback, as even the simplest systems which can be expressed in terms of rate equations include feedback.

[17]The condition for exactness of the differential form above is that both the activation-inhibition network relationships and the agonist-antagonist network relationships are completely symmetrical with respect to an interchange of any pair of indices. According to Rosen, this also is a highly nongeneric situation. It is possible to consider a hierarchy of networks, u_{ij}, u_{ijk}, u_{ijkl}, and so on, each modulating the properties of those below it in the hierarchy. If the system is describable by a set of rate-equations, every layer above these is found simply by differentiating. But, on the other hand, if any of the differential forms constructed using these quantities are inexact, the layers become independent of eachother and it is not possible to find a rate-equation formulation under any circumstances.

that the mind can directly apprehend the particular. In the nominalist position there are only general words like "dog", and no universals in the sense of entities like "doghood". The logical (if extreme) conclusion of this viewpoint has to be that there is nothing in common between the particular objects covered by the same general name. This position, which Watanabe [19, p.52] refers to as *radical nominalism* is discussed further below in section 2.3.1.

For *conceptualism*, universals are thoughts or ideas in, and constructed by, the mind. That is, universals are concepts in the mind of those who understand the general word whose meaning the universal is. Plato had rejected this notion as an explanation of why the world is as it is, which he claimed his Forms explained. These ideas came to the fore in the seventeenth and eighteenth centuries and are largely associated with empiricists such as Locke, Berkeley and Hume. Locke, writing in his "Essay Concerning Human Understanding" (quoted in [19, p.52]) claims:

> *General and universal belong, not to the real existence of things, but are inventions of the understanding, made by it for its own use, and concern only signs, whether words or ideas ... Words are general when used for signs of general ideas and so are applicable indifferently to many particular things; and ideas are general when they are set up as the representatives of many particular things. But universality belongs, not to things themselves, which are all of them particular in their existence ... [when] we quit particulars the generals that rest are only creatures of our own making, their general nature being nothing but the capacity they are put into the understanding, of signifying or representing many particulars.*

There are a number of points of interest to us here. The most straightforward one is that universals or concepts or generalisations are constructs of the mind which are found to be *useful* and can be represented as, or are equivalent to symbols (cf. Locke's use of the term "signs"). Secondly, he pointed out the role of general idea as an abstraction, leaving out all the particular ideas of individual particular objects that are not common between the objects. It is interesting to note that this is exactly the same sentiment as the modern notion of abstraction used in pattern recognition.

There is however, one point with which we would like to take issue, i.e., the notion of the real existence of particular objects. This, we argue, is itself a construct of the mind, for an object is not perceived independently of observation. Perception, we submit, is based solely on primitive observations, or measurement of predicates, *and* the relationships between measurements. Perception as such, consists only of the satisfaction of generalisations (concepts or universals). Popper makes what is essentially the same point, claiming that association psychology — the psychology of Locke, Berkeley and Hume — was merely a translation of Aristotelian subject-predicate logic into psychological terms.

> *Aristotelian logic deals with statements like 'Men are mortal'. Here are two 'terms' and a 'copula' which couples or associates them. Translate this into psychological terms and you will say that thinking consists in having the 'ideas' of man and mortality 'associated'. [44, p.76]*

The essence of the position of perception-mediated-by-the-universal is that the reality of particular objects is something that can only be inferred on the basis of relationships between observations and cannot be used as the basis for perception itself. So, while some of what Locke is saying seems plausible, we must disagree on the most fundamental point about the direct perception of the particular. Natually this position is not one that can be *proved.* While some arguments are presented here in its favour and to justify its adoption, it is primarily intended as a working assumption whose implications may be more useful than most of the alternatives.

Interpreting the general idea as a mental image, Berkeley attacked Locke's method of abstraction. He claimed that:

> *we cannot have an image of a man who is neither tall nor short, or who is simultaneously tall and short. Even if we think of a man in general, we have to imagine a man with a certain tallness. [19, p.54]*

The alternative interpretation of "general idea" as a logical condition or abstract symbol (particularly in terms of abstract concepts such as white or large) is immune to Berkeley's criticism, but also of little use in understanding perception.[18]

Berkeley's own belief on the nature of the universal has much in common with the modern ideas of epicentric pattern recognition, and Watanabe's notion of paradigmatic symbol. He claimed:

---

[18]Note that in many cases here, both the philosophical ideas presented and the counter arguments seem to both have merit, certainly to the author's mind. Some of the complication arises because of the confusion between the complementary processes of classification and grouping, in pattern recognition terminology, which are both forms of generalization or inductive inference, (see section 6.2).

that all general ideas are nothing but particular ones [ideas], annexed to a certain term which gives them a more extensive signification, and makes them recall upon occasion, other individuals which are similar to them. [19, p.54]

Hume goes further, claiming that it is impossible to conceive any quantity or quality without forming a precise notion of its degree. Everything that exists is, for Hume, a particular idea. Yet, we are not excluded from considering general concepts which are possibly ill-defined or involving an infinite extension, by the finiteness of our mental capacity. This is because

... abstract ideas [though] in themselves individual, may become general in their representation. The image in the mind is only that of a particular object, tho' the application of it in our reasoning be the same as if it were universal. When we have found a resemblance among several objects that often occur to us, we apply the same name to all of them, whatever difference we may observe in the degrees of quantity or quality, and whatever other differences may appear among them. After we have acquired a <u>custom</u> of this kind, the hearing of that name revives the idea of one of these objects, and makes the <u>imagination</u> conceive it with all its particular circumstances and proportions.
(Chapter VII, A Treatise of Human Nature), [19, p.55].

And again, from chapter IV:

All simple ideas may be separated and reunited differently by the imagination. There is no fixed rule for recombination by the imagination, but there exist certain guiding principles. These principles are to be regarded 'as a gentle force, which commonly prevails. ... nature in a manner pointing out to everyone these simple ideas, which are most proper to be united in a complex one. The qualities from which this association arises, and by which the mind is after this manner convey'd from one idea to another, are three, viz. Resemblance, Contiguity in time or place, and Cause and Effect. [19, p.56]

There are three crucial aspects to Hume's ideas: the notion that only particular examples are ever instantiated in the mind; that these particular examples can 'stand for' others in what amounts to generalisation, and finally, that the extent and type of this generalisation is determined by a process of identifying resemblance.

The first aspect of Hume's ideas is precisely what we would expect within the framework of visual perception presented here: the mental representation of any perceptual idea, either in the perception or in the recollection, consists of the simultaneous activation of many different perceptual concepts at different levels of abstraction. The latter two aspects of Hume's idea amount to a description of an inductive generalisation, in this case, based on resemblance. Now, Hume had rejected the notion of a logical induction, but here describes an extra-logical non-necessary generalisation as an intrinsic part of perception. Like Locke and Berkeley he believed that complex ideas (perceived or recollected ideas) could be constructed from simple ideas (more primitive sensations) by a process of learning through connection or association. In common with the other empiricists, he rejects innate ideas because they mean fixed association but he differs in his description of the process of learning through association. He claimed that there are no fixed rules for recombination (of simple ideas) by the mind. Rather, there exist certain guiding principles related to resemblance, continuity in time or space and cause and effect – principles which can be regarded as the non-necessary non-compelling heuristics which must be introduced in order to overcome any inductive ambiguity[19]. (Cf. Chapter 6).

From Plato, then, we have the idea of Form which emphasises properties (predicates) and function over substance[20], and also the rejection of the real existence of particular objects. This leads to the recognition that perception involves relationships between groups of observations or measurements of predicates – this is the only way we can come to know anything. Conversely, what we perceive with our senses are not objects or sense data, but simply observations based on previous experience of similar observations[21]. In truth, Plato's concern was principally with the nature of reality and what we are doing here is carrying over his ideas to a view of the nature of perception. From the conceptualists on the other hand, we carry over Hume's ideas on (an extra-logical) inductive association of particular observations to form perceptions. In addition, we get the notion that the visual sense only has a particular perception – not an abstraction with fewer predicates – but that this particular perception can be used qua a universal.

[19]Hume's "guiding principles" might be the early precursors of our modern information processing principles, like the Hebb-type rules, or Linsker's *infomax* principle, or Barlow's ideas on reliable, non-redundant communication elements.

[20]We are not particularly interested here in the source of these Forms according to Plato, or their status as "objects". It is the *use* of the Forms which is instructive in terms of developing an understanding of perception.

[21]We return to discuss the implications of the regress involved in this idea later.

It may seem trite to say this, but in modern parlance, this view is in essence that perception involves the communication of signals, and the processing of these signals so that decisions can be made on the basis of them. The nature of these decisions is not *determined* by the properties of the physical world alone (that is, the existance of objects with given properties), but by the nature of the perceptual system whose structure is compatible with a particular physical world, and whose particular decisions about what is to be perceived, are *triggered* by signals from this world. What we would like to know is, what is an appropriate architecture for such a perceptual system, and what are the information processing principles which give it its structure, consequent on past experience (particularly at critical times)? That the above discussion on objects and universals needs to be entertained at all, in order to arrive at this conclusion about signals and decisions, is indicative of the somewhat confused thinking that characterizes the representational position.

The antithesis to the position taken above on the role of the universal in the processes of perception, centres around the real existence of particular objects, which can only be seen one way – the way they objectively are (i.e. objectivism). Seeing then involves regenerating some explicit internal representation of the nature of these objects, mediated by objective (though possibly ambiguous and noisy) information about their nature (i.e. representationalism). In classical philosophy, this was effectively the nominalist position, which though anti-realist, curiously shares much with Aristotle's emphesis on substance. Watanabe shows how the purely logical extrapolation of nominalism is untenable, and it is to this subject that we turn in the next section.

2.3 Similarity and Radical Nominalism

The notion of resemblance or similarity used by Hume deserves some examination. Hume makes it clear that he does not mean that the notion of similarity is a necessary device for producing association. While the notion of similarity reflects nature, he does not claim that it is innate (and therefore fixed), nor that it is always correct. Rather, he says that it is the product of mental habit, and attributable to human nature which he does not claim to explain.

Most modern pattern recognition is based on some form of *similarity theory* – the commonsense view of classes as a collection of particular similar objects. That is

(i) what really exist are particulars, not universals;

(ii) the particular objects in a class are bound together by *similarity*.

Here, the non-necessary, non-compelling induction associated with Hume's notion of similarity is often forgotten – or replaced by a heuristic device which is not always made explicit. It is interesting to note how heavily this view of pattern recognition is based on the Aristotelian philosophy of the reality of particular objects – the idea of a fixed neutral object-predicate relationship and the existence of natural kinds. We have already mentioned some objections to these Aristotelian ideas. Further objections, some counter-examples and a thorough discussion on the ill-defined nature of the notion of similarity can be found in [19, Chap. 3].

2.3.1 Theorem of the ugly duckling

There is, however, a more serious problem with similarity theory, than objections to the reality of particular objects and the lack of a consistent definition for similarity. Even if a set of particular objects is assumed and a similarity relationship defined in a suitably restricted domain, it can be shown that there is no logical or empirical way to distinguish any pair of objects or subsets of the set of objects. Using logical connectives and deductions *only*, all objects look *equally similar.*

This somewhat startling result comes to us in the form of the quaintly titled *"Theorem of the Ugly Duckling"*[22]. This theorem states that:

> *Insofar as we use a finite set of predicates that are capable of distinguishing any two objects considered, the number of predicates shared by any two objects is constant, independent of the choice of the two objects. [19, p.82; 47, p.376]*

The use of the number of predicates shared (simultaneously affirmed or denied) as a measure of similarity does not restrict the range of validity of the theorem. This is because any given definition of similarity on a finite set of distinguishable objects can (if necessary using an arbitrary degree of quantization of continuous variables) be used as a basis for a similarity measure of this form. According to Watanabe this theorem

[22] The title seems to arise from the fact that logically, and on the basis of empirical data only, a swan and a duck on the one hand, or two swans on the other, (insofar as they are distinguishable) are equally similar.

is simply a rigorous expression and logical implication of a *radical* form of *nominalism*: there is no such thing as similarity or dissimilarity as far as a logical treatment or empirical data are concerned. Any general concept is applicable to any arbitrary set of objects.

Clearly this cannot be the whole story. The existence of all intelligent life is based on notions of similarity, classification and generalisation applied to the natural world. The ability to make and use (and possibly share) mental concepts and generalizations is probably a necessary characteristic of intelligence. Even the conditioned reflex described by Pavlov and others requires some form of generalization (presuming the ability to discriminate different stimuli). There must be some method of escape from the "booming buzzing confusion" imposed by this radical nominalism. The problem does not seem to lie with the particular objects and the definition of similarity. We can quite easily apply measures based on a suitable definition of similarity in a restricted problem domain and derive useful generalizations or classifications. The fact that serious problems arise when we try to carry over our generalizations or classification processes to logico-deductive systems, (which computers implement), indicates the source of the difficulty. The proof of the Theorem of the Ugly Duckling is based on a logical treatment of empirical data. What we need is some extra-logical or extra-evidential elements to the argument for measures of similarity. That is, we need some predicates to be more important that others or some predicates to be incompatible with (and therefore to interfere with) others. The latter case is a denial of the suitability of an Aristotelian or Boolean logical framework for a formulation of the problem of similarity. This is examined further is Chapter 6 below.

2.3.2 Axiology and radical nominalism

But we do not need to go to this extreme to defeat this form of radical nominalism. Within an Aristotelian framework it is sufficient to allow the introduction of a scale of importance of predicates such that the resulting classifiers are *useful*, or have some *value*.

> *To be similar may be to share more of the important predicates. But, how can we evaluate the scale of importance? To answer this question, we have to reflect on the reason why we use similarity and classification in life. The answer is because it is useful, In other words, our scale of importance must be such that the resulting classifications carry utility or value. We can*

> *overcome the radical nominalism only by axiological considerations. I do not hereby mean any ultimate value, but various instrumental values towards more fundamental ends. [19, p.84]*

Watanabe considers that mathematically each predicate must be assigned a different weight (what he calls a *preferential ponderation*) that varies depending on the use that is going to be made of the resulting classification. He uses an "entropy"-type function to measure the distribution of these predicate weightings, concluding that we can only see similarity, and hence grouping of objects, if there is an "uneven" emphasis on the empirical data about objects:

> *epistemology can subsist only through its interaction with axiology. It will be deprived of its major function, concept formation, if it relies only on observational experience and logical manipulation. What makes cognition possible is the evaluative ponderation, whose origin is aesthetic and emotional in the broadest sense of the term. [19, p.88]*

In the case of sensory perception of animals, inter- and intra-sense scales of importance (what Watanabe refers to as "value-orientated ponderation") must surely have been dictated by the value of individual survival. There can be no *logical* justification for such a scale of importance however. Neither can there be a justification based simply on the *empirical* data gained from testing predicates. Like the tautology in Darwinism: "the fittest are those that survive", the importance of predicates have no more justification than that ascribed by the mechanics of evolution or the aesthetics of a computer programmer.

2.4 Induction and Deduction

We have already met a number of cases of the ideas expressed by the adjectives 'non-logical' and 'inductive', being applied to processes. Examples are Hume's "gentle force which commonly prevails, ... a non-necessary, non-compelling advice" and Watanabe's "axiological" overcoming of radical nominalism. Similarly the process of paradigm-oriented pattern recognition is an inductive process, as a classification (the application of a universal concept to a possibly infinite number of objects) is made by generalization based on a finite number of examples. There have been many attempts to properly define the nature of inductive inference and to elucidate rules for the application of such inferences. We examine some of these here.

2.4.1 Induction and evidence

A definition of induction which would probably find most general agreement is one which says what induction is not: induction is the opposite process to deduction, i.e. any inference whose premises do not ***logically entail*** its conclusions. The following definitions are also often used: deduction is the derivation of particular facts from a general rule and induction is the derivation of a general rule from a finite number of particular facts. Alternatively the distinction is defined as one between logical or demonstrable inference and probabilistic and non-demonstrable inference [19, p.97]. There are problems with both of these definitions. In the latter case there does exist a probabilistic deduction and a logical refutation (called "infirmation" [44]) in induction. In the former, the emphasis on particular and general belies the real nature of induction. The following inferences illustrate some of these points. (Here *H* stands for the hypothesis or general rule assumed; *A* for an auxiliary fact and *D* for an experimental datum)

H: John is a boy.
A: John has brown hair.
D: John is a boy with brown hair.

H: All cars have wheels.
A: All wheels are round.
D: All cars have round wheels.

H: Bill is a boy.
A: Bill has brown hair.
D: Some boys have brown hair.

H: All emeralds are green.
A: Everyone has an emerald.
D: My emerald is green.

Of these four inferences [54], the first has specific premises and a specific conclusion, the second has general premises and a general conclusion, the third has specific premises and a general conclusion, and the fourth has general premises and a specific conclusion. All four are deductively valid inferences.

Now consider the following:

H: Men are mortal.
A: Socrates is a man.
D: Socrates is mortal.

D: Socrates died, Plato died, Wittgenstein died, etc.
A: They are all men.
H: Men are mortal.

H: A man is, with probability of 95%, right-handed.
A: Mr. X is a man.
D: Mr. X is, with probability of 95%, right-handed.

D: 95% of men examined were found to be right-handed.
A: They are non-biased random samples of men.
H: Men are, with 95% probability, right-handed.

Of these four inferences [19, p.98], the first is deductive, the second is inductive, the third is a probabilistic deductive inference and the fourth is a probabilistic inductive inference. The point that is being emphasised here, is that the distinction between deductive validity and inductive strength lies not in the generality or specificity of the premises and conclusions, but instead lies in the evidential relationship that exists between them.

On this point of the evidential relationship between hypothesis and conclusion Watanabe criticises Carnap's theory of induction which is based on the so-called "necessary view" of probability also associated with Keynes [19, p.102]. According to this view there is one and only one relation – called a probability relation – between any two propositions (as the premises and conclusion of a probabilistic inference). This relation is a numerical value and depends only on the connections between the two propositions and not on any human opinions or on any other factors outside these two propositions. Thus, Carnap introduced the probability $c(h,e)$ of a hypothesis h, which is a function of the relationship between the hypothesis h and the relevant evidence e only. That this view of the relationship between hypotheses and evidence is completely untenable is shown below.

2.4.2 Probabilistic interpretation of induction

One way of discussing notions of deductive and inductive inferences is to adopt the standpoint of a personalistic foundation of probability (as opposed say to the frequency interpretation[23] of the formal probability calculus, or the necessary view mentioned above). Such personal probability is applicable to reproducible as well as non-reproducible events. Here the past record of frequency reflects on the credibility distribution and the credibility distribution in turn reflects on the prediction about the future frequency. The so-called "objective" frequency view of probability tries to eliminate the role of the person who estimates the probability and makes the prediction.

In the case of inferences expressed as propositional inferences (if (*hypothesis*) and (*auxiliary fact*) are true, then the (*deduction*) follows), we need to assign probabilities to each proposition of the three categories. The probability of A is usually assigned the value one as it usually represents a fact. Then the deductive probability, which can be described as the probability of occurrence of the empirical event D based on the assumption of H and A, is given by $p(D|H \cap A)$. In the case of a logical deduction as opposed to a probabilistic induction, this probability takes one of the values zero or one. The inductive probability or credibility of a hypothesis $p(H|D \cap A)$ is the degree of confidence we place in H on the basis of D and A. Using Bayes' theorem we can rewrite this as

$$p(H|D \cap A) = p(D|H \cap A) \, . \, p(H)$$

[23] The frequency view of probability can only be used in a roundabout way to discuss induction. See [47, pp.350-351] for objections to the frequency view of probability. The personalistic view, which involves a behaviourally observable way of measuring subjective probability is also described in this reference [pp.352-361].

where $p(H)$ is the degree of confidence we place in hypothesis H without taking into account the evidential fact D and the auxiliary fact A. Thus $p(H|D\cap A)$ can be thought of as an *a posteriori* inductive probability, while the personalistic probability $p(H)$ is the *a priori* inductive probability.

Hume is often credited with proving on logical grounds that induction does not exist (see for example [44, p.51]). In fact Hume was interested in the connection between cause and effect which like the universal, he believed, was not a necessary connection, but was based on an association in the mind created by "habit". What he recognized was the same origin for the formation of universal ideas and the formation of general rules by induction, thus showing that induction has no logical foundation.

At this juncture it is probably appropriate to comment on Popper's views about induction [44, p.20]. In short, Popper believes that there is no such thing as induction. To be fair however, there are two ways in which Popper's notion of induction differs from that used here. Firstly he uses the word induction to mean logical induction, and as by definition there is no logical induction, his statement is true, even if at odds with the accepted use of the terminology. Secondly, when Popper talks about induction he is really talking about the scientific method, the acquisition of scientific knowledge and the place in this process of scientific theories. While Popper does not admit a process of indication, or proof by induction[24] he proposes what he refers to as *infirmation*. This, in fact, amounts to a logical refutation in induction: if a hypothesis states that a phenomenon will not occur and it does occur then the hypothesis is rejected. This can be clearly demonstrated with the Bayesian formula given above: if $p(D|H\cap A) = 0$ (the evidence contradicts the hypothesis) then $p(H|D\cap A) = 0$ (the hypothesis is wrong). There is no need for the personalistic *a priori* probability $p(H)$ in this process of infirmation or refutation, just as it is not relevant in constructing the deductive probability $p(D|H\cap A)$.

It is argued here however, that induction in general is not restricted to this particular process of infirmation — whatever its merits in scientific discovery. Hume recognized the process of *confirmation by positive evidence* and the work described here is based

[24]The mathematical case is a different use of the term *induction*

on the assumption that some such process operates as the basis of sensory perception. Again the Bayesian formula allows the process to be clearly demonstrated: the larger the deductive probability $p(D|H \cap A)$, then the larger the inductive (*a posteriori*) probability $p(H|D \cap A)$, if the *a priori* probability $p(H)$ remains the same. The *a priori* (in the sense of prior to the evidential fact) probability $p(H)$ does influence the evaluation of a hypothesis $p(H|D \cap A)$, but in a extra-evidential, extra-logical way. The quantitative strength of this factor can never be uniquely, logically and universally determined. To this extent there is always remaining *inductive ambiguity*. This effect is easily seen in the case of scientific discovery and theory building. If experiments are repeated, particularly in an effort to refute a theory (infirmation), and all experiments continue to support the theory within the calculated margin of error, then the effect of confirmation will tend to overwhelm the *a priori* factor even if this theory is considered particularly unlikely ($p(H)$ is small). However, insofar as the body of evidence is always of finite size we can always change the value of the *a priori* factor to counteract the evidential factor. This would be the case for example if a rival theory were proposed which was much simpler or more consistent with an existing framework. In this case $p(H_1)$ for the first theory would become smaller, and $p(H_2)$ of the newer more pleasing theory would be larger than $p(H_1)$.

In this context it may be useful to introduce to the discussion a notion about scientific theories which we draw on below in the discussion of perception. Basically, this view is that there is no such thing as truth in science. Any theory is just a useful model to correlate the facts. Zukav's direct style puts the idea very well:

> *So much for the relationship between the 'truth' of a scientific assertion and the nature of reality. There isn't any. Scientific 'truth' has nothing to do with 'the way that reality really is'. A scientific theory is 'true' if it is self-consistent and correctly correlates experience (predicts events). In short, when a scientist says that a theory is true, he means that it correctly correlates experience and therefore, it is useful. If we substitute the word 'useful' whenever we encounter the word 'true' physics appears in its proper perspective. [50, p.287]*

A similar substitution – using 'useful' for 'true' – puts induction, particularly in perception, into proper perspective. Induction is not a necessary logical progression. Any particular inductive step (or inductive jump, or decision) finds its justification not in how it was achieved, but in how useful it is. An induction is no more true (in the sense of necessarily following from reality) than any scientific theory is true. As soon as a more useful theory comes along [55] confidence in the original theory wanes.[25]

2.4.3 The necessary view of probability

Returning to the discussion on the relationship between probability and induction, the Bayesian formula above also helps to clarify the flaw in Carnap's necessary view of probability. In the notation used here, the probability of a hypothesis based on the relationship between the hypothesis and the relevant evidence e, $c(h,e)$, is given by $p(h|e)$. Now

$$p(h|e)p(h) = p(h|e) = p(e|h)p(e)$$

but before using this probability relationship it is necessary to determine which of the five probability distributions are determined by the nature of the problem. If the joint probability $p(h \cap e)$ (which completely describes the relationship between the hypothesis and evidence) is available, then all of the remaining four probability distributions are implicitly contained in it and can be directly derived by straightforward computation.

However, instead of computational processes based on the joint probability, what we usually have are inferential processes (inductive or deductive) based on the conditional probabilities. That is, only one of $p(h|e)$ or $p(e|h)$ is determined by the nature of the problem. In the case of deduction, the deductive probability, $p(e|h)$ is determined by the relationship between h and e: from a hypothesis we can make logical predictions about experimental data, without any further considerations of probability. In the case of induction however, logical predictions of hypotheses based on evidence cannot be made. The inductive probability $p(h|e)$ is given instead by $p(e|h)p(h)/p(e)$. It also, therefore, depends on the prior probability of h, $p(h)$. This prior probability is independent of the relationship between h and e. If $p(h|e)$ is determined by the

[25] The process of generating new hypotheses or theories and assigning a non-zero prior probability has been referred to as *abduction* by Peirce [55]. Little is known about the process except that certainly in the case of scientific theories it is very difficult to think of even one new hypothesis. It is possible that the process of generating new hypotheses is related to the process of analogy or the idea of metaphor. These are probably tempered by the logical consistency of the mind's theoretical structure and particularly by the aesthetic harmony between different parts. In any event the mind cannot generate all possible hypotheses arising from a particular problem structure – the choice of those actually placed for evaluation is made in a highly contingent fashion. In [30] Varela comments on the complete absence of common sense in the representational position within cognitive science. In this view the aim is to successfully represent an external world which is pre-given. Yet, according to Varela, precisely the greatest ability of cognition is to *pose* the problems to be addressed at any given moment. See the relevant quotation from Varela in section 1.3.1 above.

problem we cannot also arbitrarily determine $p(e|h)$ because variations in $p(h)$ would allow this factor to immediately contradict the given value of $p(h|e)$. All that can be determined along with $p(h|e)$, without being open to contradiction, is $p(h)$. Thus Carnap's assumption that $c(h,e)$ is determined by the necessary relationship between h and e cannot be correct. Even though the Bayesian formula is symmetric, there is a fundamental asymmetry in conditional probabilities. $c(h,e)$ should include the extra-evidential evaluation embodied in $p(h)$.

2.4.4 Abduction

The over-riding question with inductive inference is exactly what process or 'reasoning' allows the inductive jump or decision to be carried out. That is, induction is not necessary (logical), nor is it based solely on the available evidence or theoretical structure, nor is it even always correct. Yet induction demonstrably happens and is usually *useful* to the organism that makes the generalization.

One possible presumption is that there exists a set of extra logical extra-evidential guiding principles (like that of the principle of simplicity), which – though giving no guarantee of the "correctness" of an inductive conclusion – do allow a tentative evaluation of $p(h)$. This factor, which often leads to the successful progress of an inferential process, is called a *heuristic*. Whether or not organisms have a fixed set of heuristics, the products of accidents of evolution, (consistently useful heuristics help survival) is not known. Another possibility is that the mind has the ability to generate heuristics to satisfy particular requirements. We return to the topic of induction to discuss the various types of induction involved in pattern recognition in section 6.4 below. The notion of minimizing entropy as a heuristic for overcoming inductive ambiguity in pattern recognition is discussed in section 4.5.

2.5 The Purpose of Perception

The status of the assumption that it makes sense to consider such things as perceptual systems has been alluded to above. The discussion on the nature of the universal, and its role in perception, certainly assumes that this is a valid assumption. The more recent work of Maturana and Varela, and more generally the so-called deconstructivist school of western metaphysics associated with names like Heidegger and Derrida, takes a very different position to the traditional one. It emphasizes what Varela refers to as the "surplus of signification", the additional significance brought to a physical world (or enacted) by cognition, and the intrinsic circularity or *reciprocal causality* of the neuro-logic of cognition [26]. In the traditional representational view, strongly embodied within the computer/robot vision community, perception – particularly visual perception – is considered as something which allows greater control over action. The primary direction of information flow is considered to be from scene to sensory system to motor system. The picture suggested by Varela and his collegues turns this idea almost on its head. Here action is considered primarily as the control of perception, and information (devoid of the semantic connotations of representationalism) is considered to flow equally from sensorium to motorium, *and vice versa.*

> *The fundamental logic of the nervous system is that of coupling movements with a stream of sensory modulations in a circular fashion. ... the state of activity is brought about* most *typically by the organism's motions. To an important extent, behavior is the regulation of perception.*

In fact, in the sort of operational explanation appropriate for discussing the organization of a cognitive system, there will be no mention of information. The description will instead involve just circular or cooperative dynamical interactions between the sensorium and motorium within the cognitive system (which give it its properties), and the circular dynamical interaction between the motorium and sensorium, which are mediated by elements of the external physical world (and which we describe as behaviour).

> *The neuronal dynamics underlying a perceptuo-motor task is, then a network affair, a highly cooperative, two-way system, and not a sequential stage-to-stage information abstraction.*

We do not attempt here to situate our discussion solely within this alternative viewpoint, though clearly our sympathies are strongly aligned with it. Our objectives here are more of a transitional nature: promoting the transition away from the traditional picture, by examining some of its shortcomings, and attempting to describe alternatives. Our approach attempts to use the language and terminology of the traditional or conventional position and make the corrections or point out the difficulties as required. The starting point for the Maturana and Varela position is a radical and complete break with the ideas and terminology adopted heretofore, and it is sometimes difficult to see at exactly what point the traditional picture breaks down. If both the ideas presented here and the Maturana and Varela position are correct, there should be some convergence, and every attempt is made to point out the parallels, where applicable, but our aim is not simply to present arguments which support lets say, the autopoietic position.

2.5.1 The eye/camera analogy

The eye and the video camera (the so-called "electronic eye") are often the subject of a superficial comparison. It is interesting to examine the limitation of this analogy as a means of emphasising the actual role of the eye [11,56]. The functional analogy probably has its roots in the discovery of the image-forming properties of lenses, and observations by natural philosophers such as Descartes of structural similarities with the image-forming optics of the single-chambered eye. The image-forming, single-chambered eye is, however, just one of an intriguing collection of adaptations which serve as the transducers for a visual sense [57]. The compound eye, for example has similar spatial discriminatory capacity to the single-chambered eye, yet does not form any type of projected "image" in the conventional sense. Taking the collection of different "eye" types as a whole, the function of the eye becomes much clearer. It is to maximize the amount of information available from the changing optic array needed to guide the animal's or the person's actions[26].

The purpose of a camera is to produce pictures to be viewed by people. It is to reproduce as accurately as possible either a single "snap-shot" of, or a continuous stream of "snap-shots" from, some part of the optic array[27]. The camera is thus principally characterised by the quality of its optics. By comparison, the image-forming optics of the single chambered eye are of extremely poor quality. As well as optical and chromatic aberrations causing blur, there are lens and corneal aberrations which cause distortion of the image. There is also the curvature of the retina which means that straight lines are curved and metrical relations in the image do not correspond to those in the external world. Apart from the small area corresponding to the fovea which accounts for less than 1% of the retinal area, the

[26]Recall that in line with Varela's view on the role of "information" in explanation it is possible to use the term in a pedagogical or expositional way as long as we do not assume that this is operational for the system.

[27]Computers can even help in this process by making changes to visual data -- enhancing or "cleaning up" an image -- which eases the task of the viewer.

light sensitive elements are spaced in a fairly irregular pattern. The average spacing of the daylight sensitive cones changes by two orders of magnitude over the surface of the retina. Finally, the retina is overlaid by a matrix of blood vessels and there is a relatively enormous hole (the blind spot) just off-centre in each retina where the ganglion cell axons exit. The single chambered eye then, is not characterized by its image quality, but by its ability to move, to adapt, and to transmit an incredible amount of information to the brain, considering the nature of the components of which it consists.

Although all these factors cause predictable distortion of the retinal image and it might, in principle, be possible to correct for them, producing a better representation of the external world, this approach would betray a misconception of the rationale behind the function of the eye. The process of extracting information from the changing optic array has already begun in the retina. What is transmitted to the visual cortex along the optic nerve is not a "neural image". It is a highly non-redundant set of physical signal measurements[28] extracted from the visual patterns formed on the retina. In fact it is not even a transformed image in the sense of a synchronized set of data. It is a continuous flow of information, where information concerning simultaneous events on the retina can actually reach the cortex at very different times depending on the nature of events.

2.5.2 The status of functional explanations of perception

This discussion emphasises a point also made by Francis Crick [57] about the brain as a whole. We must be very careful about interpreting what any part of the brain is doing because:

> *the brain handles information in ways quite different from those we might have guessed at ... we are deceived at every level of our introspection ... our capacity for deceiving ourselves about the operation of our brain is almost limitless.*

This is a point the author believes cannot be emphasized too much. It *is* interesting to try to interpret the function of a particular part of the eye or brain. The brain however, is so vast and so complex that such interpretations must be tentative to say the least.

---

[28]Measurement is used here in the normal engineering sense rather than in the quantum mechanical sense used elsewhere in this report.

We must be careful not to claim that such and such an interpretation is *all* that is going on. There are two reasons for this. Firstly, to claim on the basis of introspection, neurological data or mathematical arguments that early biological vision processes "look like" edge detection or feature detection or Fourier analysis does not mean that one or more of these must be the first step in a computer vision implementation. That is an abuse of the role of metaphor. Marr shows that he was sensitive to this point (see e.g. [9, p.348]) but his notions of computational theory, representation and separate levels of analysis are particularly susceptible to abuses of this type.

A second, even more fundamental reason is pointed to by Varela [26,24]. This is that our descriptions of the functional role of aspects of a living system are based on our access, as observers, to both the system and its environment. But the concepts which we use to describe the interactions between the system and its *environment* are privileged to us in our role as observers – the system does not have access to them and they are not *operational* for the system. The system only has access to its *world* through its interactions with what it is not. Whatever in its world is important to it, is not so because of any objective intrinsic importance of the thing in itself (objectivism or representationalism), or because the system has "decided" in complete isolation to assign importance to it (solipsism). Instead the system and its *world* are the result of the history of the system's interactions with its *world* in the ongoing maintenance of its (the system's) identity. What in the system's world is important to it, is not necessarily important, *or even necessarily accessible*, to us in our world consisting of the system and its *environment*. We do not necessarily have access to the system's perception of its *world* (as we would if there was an objective, well-defined fixed-information world), nor does it have independent channels of access to its *world* other than through its interactions as a result of its operation [26].

Marr emphasised the role of constraints in the discovery of the appropriate computational theory for visual perception [9, p.23]. There are two problems with his use of constraints in this way. The constraints are usually formulated in a framework which encodes our biases. They depend on the pre-existence of a perceptual system, ours, without which they have no meaning. They exist because of "assumptions" made by our perceptual system in overcoming inductive ambiguity, not as a causal reason for these assumptions. This point is just a rephrasal of the objections to the

Aristotelian object-predicate notions. Because the visual system has found it useful to "code" the external world in this way, this does not mean that this is the way the external world is. It is, so to speak, a useful illusion created by the system, not the *modus operandi* of the system. A more practical problem associated with the use of constraints to discover a computational theory is that of the sufficiency of the constraints. It is very likely that in all but a few reasonably trivial cases, such as the cash register example that Marr uses [9], the computational theory will be underconstrained by the available knowledge of what the system must achieve. Thus further unjustified biases enter the problem through the introduction of heuristics to overcome the ambiguity in solution. The assumption that we can understand the direct computational theory of a perceptual system – "what is being computed, and why and that this is in some sense sufficient" is in the author's belief, a dangerous assumption which has yet to be justified.

2.5.3 An approach to understanding perception

A much more cautious, and we believe in the longer term, more profitable approach is espoused here. This approach does not presume that we know what the end-point of perception is. (The notion of an end-point to perception, such as a 3-dimension representation, may not even be well-defined). Nor does it presume to know how the so-called "output" of perception is related to action. It recognizes that all perception has to work on at any time is previous perceptions and an information processing system with suitable "dispositions". Concepts of space and time, of objects and depths and 3-D representations are not innate; nor can they be proved to be really as we conceive them. They are simply our limited conception of what our perceptual system is doing based on interpretative mechanisms provided by our perceptual system. What is real, is the coordinated activity of layers of neurons, their development and adaption, the structure of their maps, their relationships to their neighbours – and more fundamentally – why things are this way. That is, not "why?" in the sense of "top-down" and imposed computational theories and representations, but "why?" in the sense of the only thing that perceptual mechanisms deal with – data signals transduced from the external world, and their own organisation at any particular moment. This change of philosophy is summed up quite succinctly by H.B. Barlow:

> *... for detecting global, or non-local, properties a form of representation is required that brings signals of the events to one locus in the cortex where*

ever they occur at the sensory surface, and to do this non-topographical mapping is required. In this way the whole cortex acquires its unity again, for it all becomes association area, and the primary projection areas with good topographic maps are simply regions specializing in the detection of local association. This is an important change of viewpoint, for the natural question to ask about a particular cortical locus becomes 'What types of information are brought together here?' rather that 'What is represented here?' [16]

From this point of view, a knowledge of the constraints or limiting factors on a real perceptual system (rather than its imagined computational equivalent) convey much more about the nature of the information processing problem. We shall see that very interesting inferences can be drawn from such things as the regularity of receptor spacing in the fovea and periphery of the retina; the number of ganglion cells compared to the number of receptors; the information capacity of neurons; the connectivity of the cortex and so on. Fundamental to all of this must be the following realization: Regardless of the length of time over which a species has evolved to its current form, its perceptual system is not simply an ad hoc assemblage of components with suitable properties engineered to implement a computational theory which has been found to be useful. There simply are not enough variables in genes, nor enough generations in the evolutionary time-span to discover and specify arbitrary perceptual systems and brains in this way. Rather there must be underlying information-processing principles which have been discovered by evolution and which can be exploited to a greater or lesser advantage by each individual species depending on its biochemical makeup and its ecological niche. These information processing principles rather than a high level computational theory are what would be common across species and across perceptual modalities. These are the goals of our research – to recognise the importance of their existance and the nature of their operation.

In 1979, Francis Crick [57], one of the discoverers of DNA, warned of two dangers which must be avoided in constructing a general theory of the brain. One is the fallacy of the homunculus which is discussed below, the other is the fallacy of the "overwise" neuron. If a computational theory is claimed to be any more than a "commentary" on an evolved information processing system, it has fallen foul of one aspect of this latter danger. In an attempt to understand and reproduce perceptual systems, such as a visual sense, or particular modalities such as stereopsis, we should not presume to know or be able to discover all there is to know about sensing in that particular way. Our guesses and our framework are hopelessly biased. Instead we should try to discover the constraints on existing visual systems and their underlying information processing principles. We should try to discover what information is being brought together at a locus in the system, taking care not to attribute too much to capabilities of any particular components or subsystems. It is with these ideas in mind that we examine information processing in the retina and cortex in the following chapters.

2.6 Summary

The ideas of object-predicate inversion introduced by Watanabe are extremely useful for pointing out an alternative way of addressing the perception of an external world. In particular we have tried to argue for the view that objects are not directly perceived but rather are artifacts of our perceptual mechanisms which are mediated by the observation, measurement or evaluation of predicates. This in turn leads to the assertion that the nature of perception is one of relationships between predicates which is discussed further in chapter 7 below. In addition we see that the only way of overcoming the radical nominalism inherent in a completely logical empirical approach to pattern recognition is to in some sense associate a value or level of importance to predicates and this notion of value or usefulness is one to which we return again. Finally in a discussion on the nature of the camera/eye analogy the fallacies of the homunculus and the overwise neuron were implicated in some of the misconceptions about the role of the visual system.

Chapter 3

3 Natural Visual Perception: The Retina

3.1 Introduction

Just over a century ago Ramon y Cajal published his first studies on the retina using the Golgi method. Because of its orderly organization in alternate layers of cell bodies and intercellular contacts and the easy identification of the main direction of the nervous message flow, Cajal was able to draw important conclusions about the basic organizational principles of the nervous system. These led directly to his theories of the neuron doctrine and "dynamic polarization" of nerve cells which are the basis of modern understanding of the information processing function of the neuron [58]. Over the intervening 100 years, the retina (which developmentally is part of the brain) has continued to be a focus of great interest in terms of its anatomy, its physiology and the processing functions which it contrives to carry out on the incoming visual data. Much of the earlier work on electrophysiology (directly monitoring the activity of single nerves) was carried out on the compound eyes of invertebrates, because of the difficulty of using similar measurements on vertebrate eyes. Lately many of the gaps have been filled in with work on the eyes of higher animals including cats and monkeys and, as we shall see, the basic information-processing functions carry over with only minor modifications and extensions.

The relevance here of this material is that because the structure and function of the retina are quite well understood, and quite well explained in terms of signal and information theoretic considerations, we are given an insight into the constraints that these early parts of the visual system need to satisfy and the information processing principles that they might use to satisfy them. The importance of a thorough understanding of the concepts that underlie the operation of any part of a neural system cannot be over-emphasised. Otherwise any attempts to model these systems amounts to little more than mimicry. An example of the type of problem is the arguments used by Marr [9, 59] and others to justify a theory of "edge-detection" on the basis of what seemed to be the function of the ganglion cells in the cat retina. This theory ignored an amount of relevant data which was available even then. This data indicated that edge-detection was at the very least an over-simplification and most probably an over-extension of what could be inferred from the evidence (cf. Barlow's "overwise" neuron fallacy mentioned in section 2.5.3). Nevertheless, the same arguments continued to be used in the computer vision community for several years afterwards.

This does not mean that we should be complacent about the information theoretic explanations discussed here. After all, our explanations of the functioning of something like a neural system are just that: *our* explanations, from the privileged position we occupy as observers. They are not operational for the system. Thus, attributing the property of detecting "edges" to a system is dependent on our ability to generate and discriminate the concept of an edge, and simultaneously describe the operation of the system with our criteria of discrimination – they are not properties that are theory neutral. On the other hand, information theoretic ideas are at least nearer to the operational level of a system than explanations in terms of edges or features, and are therefore less tied-up with our particular conception of the way we see the world (our *world*). These ideas are explored with an emphasis on the constraints that are responsible for the present form of the retinal structures being examined.

3.2 Retinal Structure

One of the first and most striking discoveries was made by Hartline and Graham in 1932, recording from the receptors of the compound eye of the horseshoe crab Limulus [56, p.34]. They found that the receptor generated an output signal which was approximately proportional to the *logarithm*[1] of the incident light intensity [60]. This logarithmic compression allows Limulus to see over a 6-7 order-of-magnitude range in input intensity (26 octaves) with neurons which have a dynamic range of 2-3 octaves. It was also found that the receptors' output activity adapts from an initial peak value at the onset of a stimulus to a sustained value after approximately one second – both peak and sustained values being related to the *logarithm* of the intensity. Thus,

[1]There has been some debate as to whether there are theoretical reasons why a log compression should be used and if the compressive process in biological vision systems is best approximated by a log. It seems that in any case, the actual function is different for different species, and different light levels, and is often best modelled as a power law [60, 11].

slow temporal changes in light intensity such as those due to diurnal fluctuations, are filtered out and cannot be seen.

Similarly Hartline *et al* discovered in 1956 [56, p.36] that slow spatial change in light pattern across the eye are also ignored. The mechanisms which are mainly responsible for implementing these filtering processes are mutual lateral inhibition (effectively high-pass spatial filtering) and self-inhibition (effectively high-pass temporal filtering).

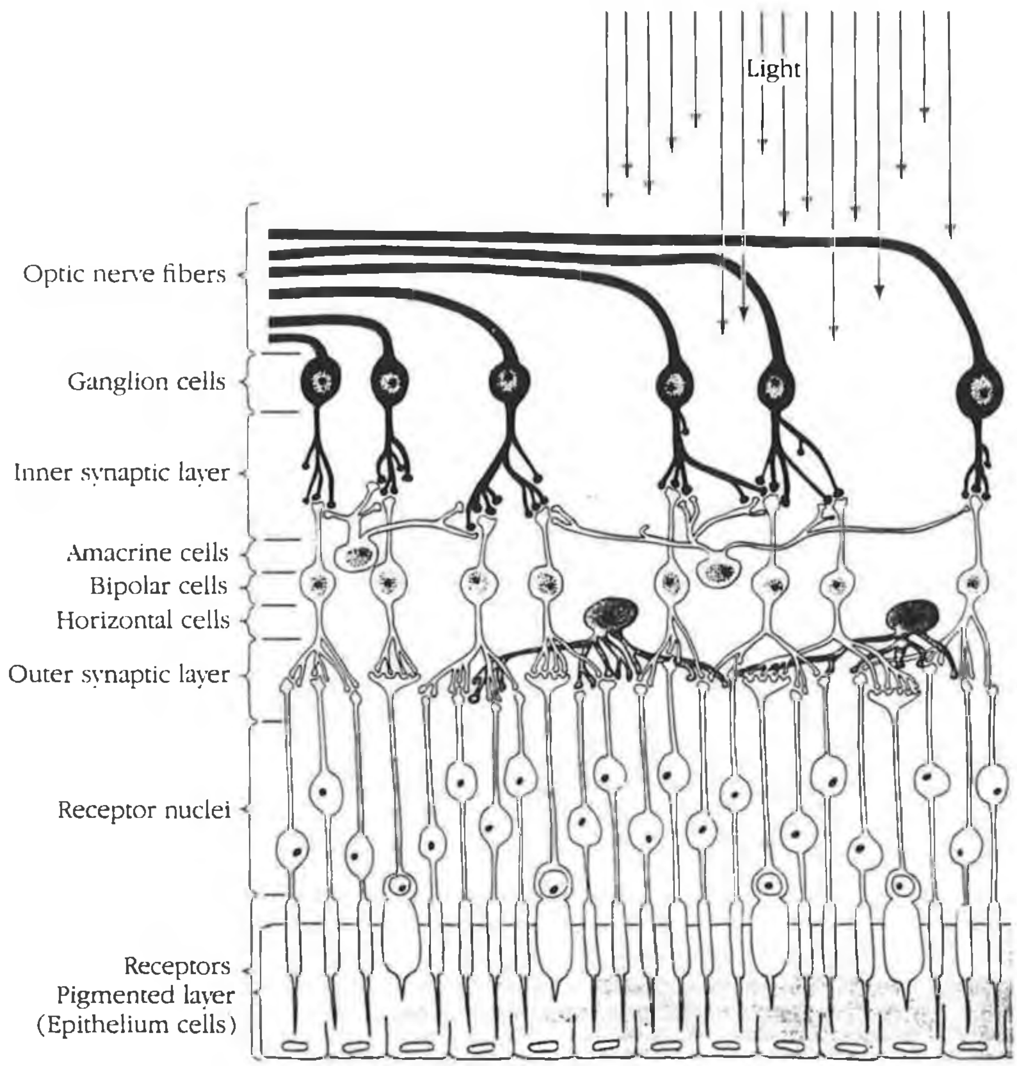


Figure 2. A highly schematic cross-section through the vertebrate retina. Note that the light passes through the semi-transparent layers of neurons before being detected by the photoreceptors. Adapted from [9, p.338].

The vertebrate retina is more complex than the arrangement of neurons in the ommatidia of compound eyes of invertebrates [56, p.39, 61]. In information processing terms however it still has much in common with these. It generally consists of three layers of nerve-cell bodies separated by two layers containing synapses made by the axons and dendrites of these cells. The first cell layer, nearest the back of the eye consists of the light receptors – the *rods* and the *cones*. The middle layer contains the *bipolar* cells, the *horizontal* cells and *amacrine* cells. The bipolar cells receive input from the receptors and feed their output to the third layer of cells – the *ganglion* cells which output directly to the brain. The bipolars are part of what is described as the direct path, linking receptors to the output of the retina produced by the ganglion cells. As well as this direct path of information flow, there are two lateral pathways. In the *outer plexiform layer* the dendrites of horizontal cells (which, usually, have no axons), form synapses with receptors and bipolars cells and "gap-junctions" with each other. In the layer of connections nearest the front of the eye, the *inner plexiform layer*, the processes of the amacrine cells, interconnect with the bipolar axons and the ganglion cell dendrites. Some ganglion cells receive their input from bipolar and amacrines, some receive input only from amacrines. The axons of the ganglion cells pass across the inner (front) surface of the retina, collect in a bundle at the optic disk (or blind spot) and exit as the optic nerve to the *lateral geniculate nucleus* (LGN) of the brain.

Probably the most striking facts about the eye are: the number and arrangement of photoreceptors; their ability to respond to single quanta of light; and their ability to resolve to the diffraction limit. There are about 120 million rods which are responsible for vision in dim light and about 6.5 million cones which do not respond in dim light but are responsible for our ability to see colour, motion and fine detail in "normal" daylight. Our principal interest here is in the vision mediated by these cones. Recent work has clarified the relationship between the cone mosaic in fovea and periphery, the eye's optics, and neural processing in the retina, in the provision of the acuity measured by experiments in visual psychophysics [62]

3.3 Fovea

3.3.1 Foveal acuity

Diffraction at the pupil and aberrations in the optics of the eye cause the retinal image to be blurred. Under optimal conditions the point spread function has a diameter of about 1 min. of arc. In terms of spatial frequency, the contrast of a retinal image falls

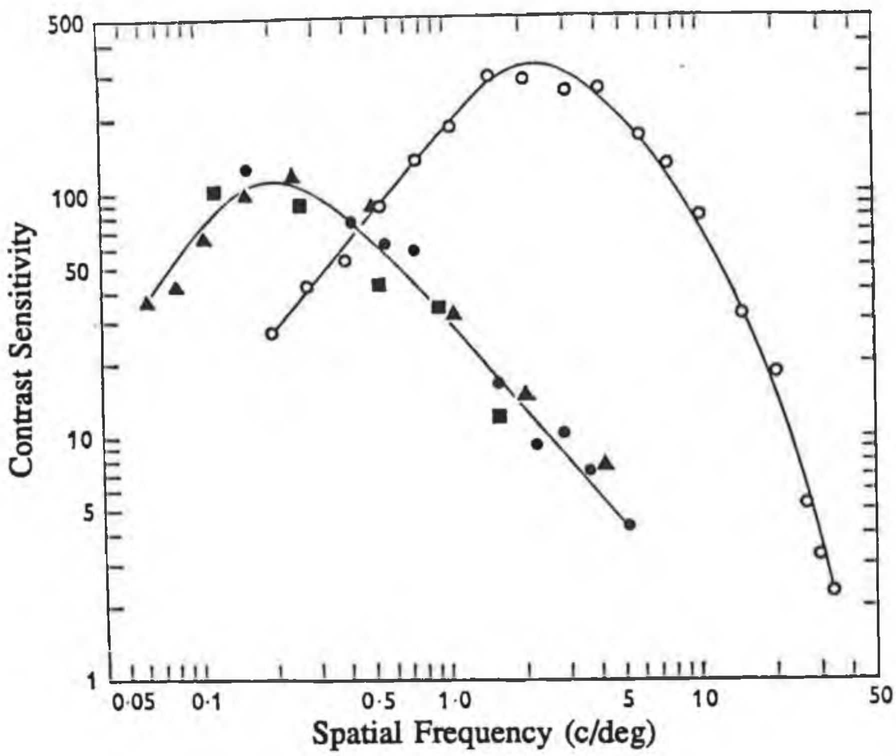


Figure 3. Contrast sensitivity curves for the cat (left) and man (right). Adapted from [11, p.118].

by an order of magnitude between 0 and 35 cycles per degree (c/°) and is negligible above 60c/°. In the fovea, the blurred retinal image is sampled by an array of cones which form an accurate triangular lattice. The cone inner segments (the active sensing element) fill most of the area between them. This normally has the dual effect of maximizing the quantum catch, (thereby minimizing the photon noise), but also of reducing the spatial frequency response because of integration of light across the cone aperture. However, with a point spread function of the eye of about 5μm, the 2.3μm diameter of the cone aperture means that the blurring due to integration across the aperture is small. (Loss of contrast is less than 25% at 60c/°). The minimum centre-to-centre spacing between human foveal cones is about 2.8-3.0μm. With a row spacing in the triangular[2] lattice of 2.4-2.6μm (0.5-0.54 min. of arc) the spatial sampling frequency yields a Nyquist limit of 56-60c/° [62]. These measurements have been confirmed by the generation of fine laser interference fringes directly on the retina of human observers who report the Moiré patterns formed by the cone mosaic. The Moiré patterns are coarsest at a frequency of 110-120c/° and can be seen up to

[2]The row spacing is important here because the lowest spatial frequency gratings that can produce aliasing in a triangular array are those oriented in one of the three orientations lying parallel to rows of sampling elements.

frequencies of 150-160c/°. They also show a 60° rotational periodicity and some distortion consistent with a slightly irregular triangular lattice.

The fact the Moiré fringes cannot be seen in the fovea under normal viewing condition indicates that the eye's optics remove the higher spatial frequencies from the retinal images which would cause aliasing. This would suggest that diffraction and aberration are the limiting factors on spatial acuity rather than receptor density. The particular form of the eye represents a choice between a larger pupil diameter, where diffraction would be reduced at the expense of aberrations, or a smaller diameter which would reduce aberrations but cause diffraction to limit acuity. Interestingly, it may actually be the cones' size and spacing which are the ultimate limit to spatial acuity, with the optics evolving to a stage where they do not interfere with the acuity achievable with the minimum cone spacing. Cones smaller that 2μm have never been found, even in very small eyes, and there is remarkable constancy in cone diameter over a large range of eye sizes [16, 63]. The apparent minimum "allowed" spacing of cones, may be required to prevent optical "cross-talk" between the outer segments of receptors. These segments are essentially short optical waveguides, with limited ability to retain incident light due to the limited refractive index differences in the biochemical materials available. Photons captured by one cone could easily cause a photopolymerization (the first step in the detection of a photon) in a neighbouring cone, if the spacing were sufficiently small. Many structural details of the eye, including the pupil diameter and optical quality, may be a consequence of the need to optically isolate cone outer segments, rather than the cone spacing being a consequence of the lack of further evolutionary pressure from the poor optical quality. The match between the highest frequency passed by the eye's optics and the resolution limit set by the spacing between cones at the foveal centre has long been known [64]. This recent explanation of which of the two factors, optics or cone separation, is responsible for the actual acuity value puts the meaning of acuity in new light [16].

3.3.2 Information processing implications of acuity

Regardless of the particular interpretation of the limiting factor or the eye's acuity, what *is* clear is that evolutionary pressure has conspired to produce in primates, an eye which stretches physical and biochemical processes almost to their theoretical limits. Few eyes have higher resolving power, with the exception of some birds of prey such as the hawk.[3] Yet, even the compound eyes common in many insects have achieved similar discriminatory capacity [11, 65]. It seems that remarkably simple neuronal systems can derive substantial survival benefit from the vast amount of data, or the high level of interactability, which high acuity vision supplies. It is mentioned above that the optical quality of the single chambered eye is poor in relation to a simple single lens camera. In an optical system where the sensing element spacing is consistently pushed to its biochemical limit, not only is the processing system (the retina and the brain) well able to deal with the vast amount of data, but it is also able to extract useful information from poor optical- and geometrical-quality image projections. This fact says quite a bit about neural systems. It indicates that it is relatively easy to generate the appropriate neural hardware and get it to interconnect and adapt to carry out useful functions. It tells us that while there may be some specialist adaptation of neural systems, as in the retina, that largely we should expect that neural systems of great power can be constructed with networks of basic neural components coupled with general information processing principles widely and flexibly applied. We shall see that with the exception of the retina itself, most of the information processing apparatus of the cortex seems to be very flexible and based on a small number of general information processing concepts. What innate structure there is, seems to specify mostly the gross mapping from site to site. This type of mapping can and has been subverted [66], exposing the effects of the underlying information processing principles[4].

As mentioned above, we do have to be careful about using evolution as a prop for our theories. Crick warns of the danger of arguing in anything other than the broadest

---

[3]The foveal depression (or pit), where the upper layers are spread aside, exposing the underlying receptors to direct light may, in the hawk, act as a further optical element. The tissue of the retina itself, which contains mainly transparent neurons and their processes, has a higher refractive index than the overlying vitreous humour. This, coupled with the shape of the interface between the two, means that they act as a diverging lens. It has the effect of both magnifying the image incident on the cone active elements and correcting for aberration. This beautiful adaptation which helps to give the hawk 2-2½ times better acuity than man, must surely represent the absolute limit of what is achievable with the available physical and biochemical processes.

[4]It has been shown [66] that early developmental manipulations of the cortex can induce sensory projections from one modality (eg. the retina) to project to cortex belonging to a different sensory modality (eg. the auditory cortex). The fascinating result of these manipulations is that they can have a significant influence on the internal connectivity, or microcircuitry of the cortex. This gives rise, eg., in the auditory cortex, to cells with response characteristics similar to visual cortex. The laminar character and interlaminar patterns of connectivity were not significantly affected — they were very much as they would have been in the visual cortex — yet these structures are not observed in the auditory cortex under normal conditions.

possible terms about constraints imposed by evolution, – whether it could have done this, or could not have done that [57]. On the other hand, The author would also disagree with the argument commonly used in computer vision (see e.g. Marr [9, p.19]) that because such-and-such is a mathematically optimal solution to a problem and evolution has had an enormously long time to work, it must have happened on this solution. If as Crick suggests, constraints are searched for by us with caution and confirmed with direct experiment then we may find valuable pointers to what is or is not possible and avoid the pitfalls of over-specificity or over-presumption. The "cross-talk" between receptors pointed to by Barlow is one such constraint which can give a perspective on another aspect of the problem (in this case neural information processing) and show the direction of a fruitful research framework.

That the neural processing of visual signals is not a factor which strongly limits acuity has been implied by further results derived with the interference fringe techniques mentioned above. These involve the measurement of foveal contrast sensitivity when the fringes are constructed directly on the retina [67]. The results suggest that neural blurring is roughly comparable to optical blurring under optimal condition. On average, only 8% contrast was required by observers to detect fringes with a spatial frequency of 60c/°. This indicates that the retina manages to carry out processing and coding functions which involve massive data compression and extended lateral interactions without substantially "blurring" spatial information – a further sign that it has not reached the limit of what is possible with this type of neural processing. These results are also consistent with the belief that the receptive-field centres of some ganglion cells are fed by a single cone. Such a belief is quite tenable as it is the simplest configuration which would maintain acuity, but it does not tell us very much about the nature of the processing carried out by the configuration.

3.3.3 Colour transduction

Much more interesting, is the way the retina has adapted to accommodate colour vision [62]. The retina contains three subpopulations of cones, each sensitive to incident light with frequencies in one of three particular bands in the visible colour spectrum. Recall that a TV camera takes colour images by separately sampling three different filtered images at the same high resolution that would be required for monochrome viewing. The retina has adopted a number of strategies which allow

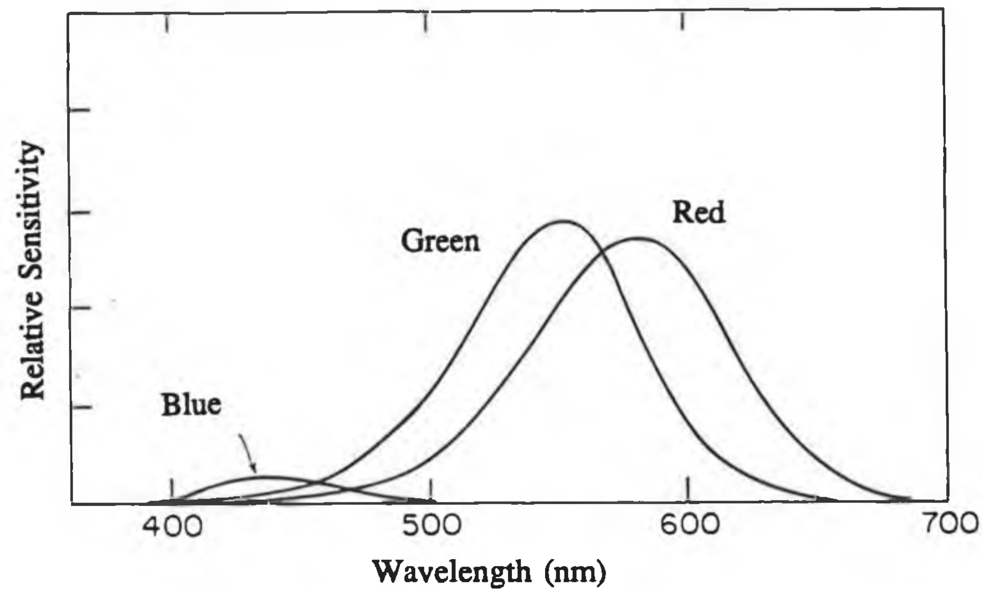


Figure 4. Typical spectral absorption curves of pigments in human retinal cones. Adapted from [60, p.32].

satisfactory colour perception without significantly degrading monochrome acuity. The combined mosaic of red- and green-sensitive cones comprise 90% of the cone population and mediate high-resolution vision. The substantial overlap in the absorption spectra of the red and green sensitive cones, coupled with the relatively smooth reflectance spectra of images of real scenes, means that the outputs of the red and green cones are strongly correlated. This makes it possible for the red and green outputs to be effectively combined to produce luminance information. This luminance data is more finely sampled than the sampling any one of the populations taken individually could produce. It is degraded little by the offset in the spectral sensitivities of the red- and green-sensitive cones. Colour information at a somewhat lower resolution is still available as a difference signal between the interspersed red and green populations of cones. As well as being at lower resolution it will be somewhat noisier than the luminance information because of the smaller differences involved between the red and green sensitivities.

In contrast to the dual function of red and green cones, the blue-sensitive cones which comprise the remaining 10% of the cone population contribute little to spatial vision, but extensively to colour perception. Their absorption spectrum is strongly shifted from that of the red and green and thus provides a strong colour difference signal. Blue-sensitive cones are particularly sparse in the very centre of the fovea, which minimizes the loss in resolution that would be caused by the interruption of the red-green mosaic. It might be expected that the paucity of blue-sensitive cones in the fovea would cause aliasing of blue components of incident light. Aliasing can be demonstrated in the lab using the interference techniques mentioned above, but there are two reasons why it is avoided in the natural course of events. Firstly chromatic aberrations of the blue components cause strong blurring of these components without substantial deleterious effects on the spatial acuity mediated by the red/green combination. Of course this also contributes to the low spatial resolution of colour perception (maximum acuity is 6-10c/° [68]). Nevertheless the brain seems to be able to combine the high resolution luminance information with this low resolution colour information to produce a unified high resolution spatial colour percept which is accurate in most cases.[5] The second way in which the blue mosaic is protected from inaccurate percepts due to aliasing arises from the fact that blue-sensitive cone mosaic forms a lattice which is somewhere between being perfectly regular and perfectly random. The net effect of this is that frequencies above the nominal Nyquist limits (given by the average inter-cone spacing) are not converted into conspicuous Moire patterns, but instead are scattered into broadband noise [62,68,69].

3.4 Extrafovea

3.4.1 Extrafoveal sampling

Outside the fovea the average spacing between cones increases rapidly. At 4° of eccentricity the average sampling rate has dropped by a factor of three. By 10° the spacing is an order of magnitude down on the foveal value. Optical quality on the other hand, declines slowly, suggesting that aliasing may be an important factor in the visual performance of the extra fovea. Apart from change in average cone spacing there is a striking change in the cone mosaic with increasing eccentricity from the fovea. By 2-3° the cone mosaic has "degenerated" into an almost random lattice with no obvious regularity and with the space between the cones filled with rod receptors. The lattice is not completely random however. A perfectly random (2-dimensional Poisson) array would have no Nyquist limit in the sense in which a regular array does, and therefore it would never produce Moiré effects due to aliasing. The cost is a

[5]Two cases where the eye is "fooled" by high resolution colour patterns are the herringbone pattern of coloured tweed viewed at a certain distance and the "pointillistic" style of painting. Here the eye can distinguish the spatial variation of luminance, but the colours of the individual spatial features cannot be resolved and "run" or bleed into each other.

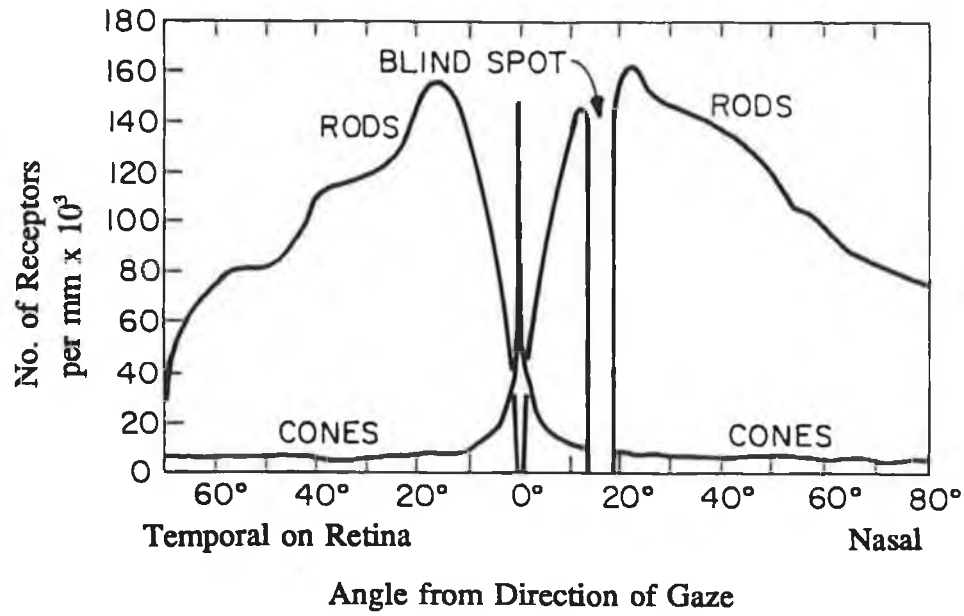


Figure 5. Rod and cone distribution as a function of angular position on the retina. Adapted from [60, p.31].

scattering of spectral energy of all frequencies into a "veil of white noise" [69]. Yellott has shown, by taking the optical power spectra of cone array sampling points, that the cones provide a novel form of optimal spatial sampling. The semi-irregular cone array is optimal in the sense that the minimal noise is introduced for spatial frequencies below the nominal Nyquist limits (computed on the basis of a regular array with the same local sample spacing or sample point density). For spatial frequencies above the local Nyquist limits, conspicuous Moire patterns are avoided by scattering the pattern energy into broadband noise [69].[6]

There are conflicting views about whether the irregularity of the extra-foveal cone mosaic is a device which has been selected for by evolution to defeat aliasing distortion [69], or if it is simply a residual disorder as a result of a lack of selection

[6]In the same way that blue cones within the fovea avoid aliasing by virtue of forming a semi-irregular lattice structure, the coloured fringes which appear on a television because of aliasing of the colour signal could be avoided. This would require the resampling of the colour signal (which is initially sampled at the same resolution as the luminance signal because of the use of RGB filtering) in a low-resolution semi-irregular way as opposed to the low-resolution regular way now used. The properties of lattices other than square and rectangular is a fascinating topic which deserves further study. For example it has been recently shown that Penrose lattices with a 5-fold symmetry have the property of blocking out some frequency bands entirely while allowing others.

pressure for further regularity [70]. Whichever is nearer to the truth there are a number of reasons why aliased patterns (for a regular array) or aliasing noise (for an irregular array) may not be a significant problem for extrafoveal vision. Most of the time the eye is not well accommodated to objects projected on the periphery: to scrutinize something, the eye makes a saccade to allow the image to be projected onto the fovea and then accommodates to bring the foveal image into focus. It is also suspected that light scatter in pre-receptoral layers of the peripheral retina causes a blurring of the projected image in these regions. Finally it has been shown [71] that natural scenes have most of their power appearing at low frequencies. These three factors all have the effect of giving low contrast at higher spatial frequencies and consequently reducing problems due to aliasing noise.

3.4.2 Extrafoveal function

A more important factor in the development of the extrafoveal cone mosaic may be the function of the periphery. The primary function of the peripheral retina in daylight is the detection of objects – particularly the detection of any type of motion. An eye/head fixation movement then allows the fovea to scrutinize the region where something was detected. This function requires the periphery to maximize contrast at sub-Nyquist frequencies, even at the expense of aliasing noise. Good contrast means that the extrafoveal cones must have good quantum efficiency and therefore need to be larger than foveal cones. Even though aliasing could be avoided by reducing the extrafoveal cone size to that of foveal cones and then reducing the sample spacing, this biologically possible step is not taken. It seems that in the case of the extrafoveal retina, a premium is placed on detection, with several factors conspiring to reduce the cost of this in terms of aliasing noise. In fact, the close match between optical quality and sample spacing in the fovea may be the exception rather than the rule. There is widespread undersampling of the cone mosaic of many other species of animals which do not have the specialized scrutinizing parvo system of primates.

The decrease in sample spacing of the foveal mosaic to the limit allowed by the refractive indices of available biochemical materials may have been closely associated with the development of the parvo system in primates. The much older (in evolutionary terms) magno system seems to be responsible in primates for the detection and processing of motion, stereo and depth cues and for gestalt grouping effects. It is not capable of the detailed and sustained scrutiny of patterns or objects necessary for fine manipulation. The newer, and in the primate, much more extensive parvo system is capable of sustained and detailed pattern scrutiny (though this ability is substantially degraded by motion and lacks the gestalt abilities characteristic of the magno system). It is impossible to say at this stage if any one of the three adaptations of (i) fine cone sampling for high acuity (ii) the parvo system for object/pattern scrutiny or (iii) the development of dexterous manipulation abilities, provided the impetus for the development of the others. It does nevertheless seem that they are closely related in some manner.

In contrast to the fovea where post-receptor neural processing hardly effects acuity at all, neural mechanisms in the peripheral retina impose substantial limitations on visual acuity. Extrafoveal acuity falls well below the cone mosaic's average Nyquist limit and this is reflected in the 5-to-1 ratio of cones to ganglion cells in the far periphery [62]. This decrease in acuity caused by the subsequent neural processing may not be surprising in view of the acknowledged role of the periphery in detection rather than examination, and the fact that the processing in the fovea demonstrates that acuity can be maintained, if required. It also further reinforces the rejection of the notion of a retina which simply transduces and codes "images". Clearly the retina carries out the role of detecting appropriate information available in the optic array. Our interest is in knowing what the "appropriate information" is, and how it can be extracted or made explicit. The neural processing of the retina and how it develops may give clues which would help to answer these questions.

3.5 Receptive Fields and Point Images

3.5.1 Receptive fields and feature-detection theories

The ideas of receptive fields and point images are important links between the activity of a single neuron and overall processing function of the eye and brain. Receptive fields in particular, provide a qualitative relationship between a cell early in the visual pathway and the sampled retinal image, and have been important in attempts to interpret the information processing function of the visual system. The receptive field of a cell in the visual pathway is the set of all points on the retina which contributes in some way to the activity of that cell. It is a concept which arises naturally from what has been the dominant experimental technique of electrophysiology to date — the

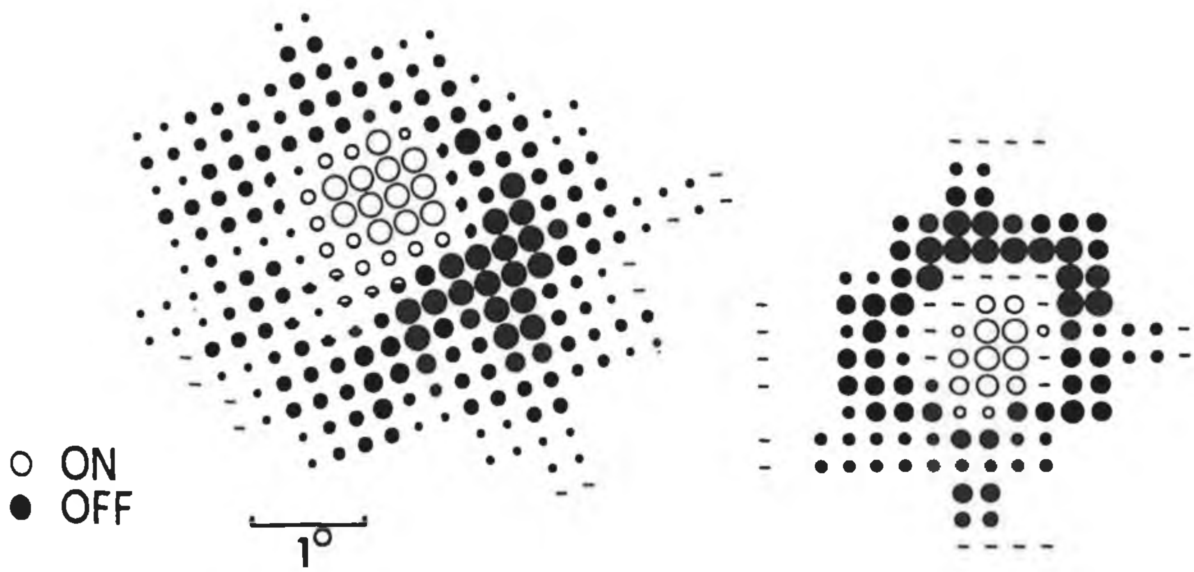


Figure 6. Plot of cat receptive fields for cells with ON-centre responses. Adapted from [11, p.105].

recording of the electrical activity of single neurons. Kuffler working in the early 1950's was the first to realize that localized spatial or temporal changes in light intensity on the retina were required in order to affect the activity of mammalian ganglion cells [72]. Recording from the retina of a cat, he discovered concentric fields on the retina, within which spots of light would consistently increase or decrease the activity of the cell being recorded. He found that there were two interspersed populations of ganglion cells. One set, called *"ON-centre"* responded with a burst of impulses to the onset of a spot of light in the centre of its receptive field or to the switching off of a spot of light in the surround area of the field. They gave no change in response to stimulation outside this concentric region. The second population, referred to as *"OFF-centre"* responded in the opposite fashion.

Enroth-Cugell and Robson (1966) were further able to differentiate two sub-categories of the mammalian cells, in terms of whether they had a linear response (X-cells) or a non-liner response (Y-cells) to sinusoidal gratings [73]. These, and further related results, particularly the interpretation by Hubel and Wiesel [74] of the functions of cells in the early mammalian visual cortex, have prompted what is the dominant theme in studies of sensory systems over the last 20 years. This is that the activity of individual neurons can represent particular aspects of a stimulus [75]. It is a view that has also strongly permeated the computer vision community, providing justification for processes of edge and feature detection as descriptions of the function of early vision. Marr's raw primal sketch with its blobs and other primitives is a more sophisticated extension of this idea [9].

However, not all observations on sensory neurons are easily reconciled with this view. Recall Crick's caution about the "overwise" neuron. He summed the idea up in terms of edge detection as follows:

> *A single 'edge detector' does not really tell us that an edge is there. What it is detecting is, loosely speaking 'edginess' in the visual input, that is a particular type of nonuniformity in the retinal image that might be produced by many different objects. ... the information relayed by a single neuron is bound to be ambiguous ... we can extract more information by comparing the firing of one neuron with that of one or more other neurons. [57]*

Recent observations support the idea that sensory coding may also depend on patterns of activity within large populations of neurons [75,76,77,78]. One example is the representation of movement in terms of a neuronal population vector. Another is the observation of correlation or synchronization between neurons in the cat visual cortex which share some attribute represented in the incident image. This includes things like having overlapping receptive fields, or having non-overlapping fields but responding to the same extended object. Even the concept of receptive field itself which provided the initial motivation for the notion of neurons as dedicated to a particular task (like feature detection), is now difficult to reconcile with this notion. Receptive fields appear to increase in size with distance along the visual pathway despite the fact that the behaviour mediated by these neurons often require a high degree of spatial resolution [75]. An idea which has recently prompted a fresh look at this phenomenon in terms of the collective activity of neural populations is that of "point-images" in the presence of retinotopic mapping. That is, the problem is examined from the point of view of how such populations deal with information from a single visual point (turning the receptive field notion on its head).

3.5.2 Point image and retinotopic maps

The original definition of *point-image* attributed to Fischer [79,80] is that it is the distribution of excitation among retinal ganglion cells resulting from illumination of a single photoreceptor. More abstractly it could be defined as the set of elements that may potentially be affected by a stimulus at some point because that point is common

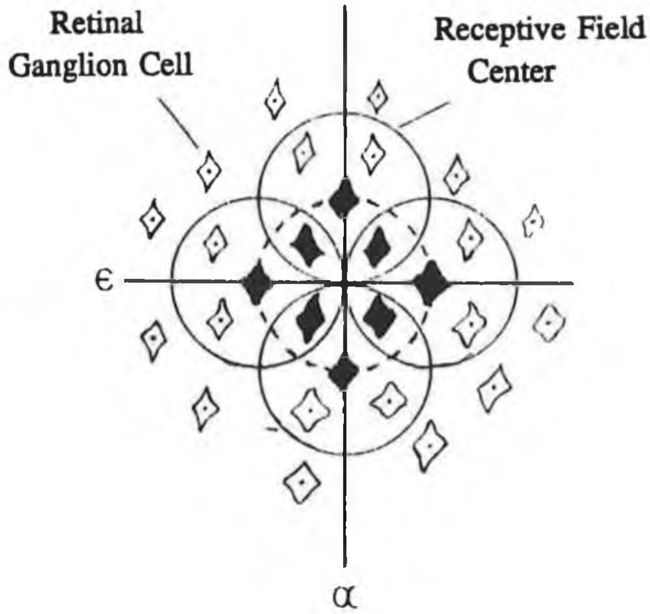


Figure 7. Relationship between receptive field centre size, cell density and retinal coverage factor. The point at the centre of the diagram lies within the receptive fields of the cells marked in black. Adapted from [75].

to all their receptive fields. Fischer's original proposal was that any point on the retina falls within the receptive field centres of a constant number of retinal ganglion cells. If we consider the fact that only ganglion cells within one centre radius of a point will cover that point with their receptive field centres, then the number of these cells is given by the product of the area of the receptive field's centre and the ganglion cell density for that particular part of the retina. Fischer's point is that even though these factors separately vary enormously with eccentricity across the retina, their product (called the *"covering factor"*) is approximately constant, because ganglion cell density decreases and receptive field size increase with eccentricity.

Peichl and Wassle, working on the cat, showed [81] that for Y-cell, the coverage factor is 3-6 for any retinal point. For X-ganglion cells the covering factor ranges from 7-10 in the periphery to above 30 in the fovea. The notions of covering factor and point image, though related, do not correspond exactly. For example, the covering factor does not take receptive field surrounds into account. Nor is the covering factor as generally applicable as the idea of point image because it is difficult to apply to cell layers other that the ganglion cells, such as those in the striate cortex. Also the diameter of the point image can increase with increasing receptive field size across the retina, regardless of covering factor.

The calculation of point images can be simplified using receptive field images. These are the projections of a receptive field in visual space (on the retina) to the retinotopic coordinate system (map) of the layer in which the cell is located. Then cells whose receptive field images contain a particular map point have receptive fields that contain the corresponding reference point in visual space.

In the cat superior colliculus, the point image of a retinal point is approximately 3mm in diameter. This clearly rules out traditional ideas of point-to-point mapping from the retina to parts of the brain in the nature of a recognizable image. After only 4 or 5 layers of synapses, the information provided by a single receptor is capable of affecting the activity of an enormous number of neurons. In the case of the superior colliculus, this type of diffuse mapping to an extended distributed representation, may be part of a mechanism for translating visual coordinates into a motor code [75].

In the pathway from the retina, via the lateral geniculate nucleus (LGN) to the visual cortex, the mapping seems to be much less regular than in the superior colliculus pathway. Here there is substantial scatter in visual space of the receptive fields of neighbouring cells. This could mean that the receptive field images are not centred on the cells, or that the retinotopic map is discontinuous or even that there is more than one retinotopic map into a number of interspersed neural populations. Again in the striate cortex of the cat, Albus showed that the product of local retinotopic map magnification and the aggregate receptive field diameter is equal to a constant of about 3mm in the area of the cortex representing the central 10° of the visual fields. This means that each retinal point provides input to the same number of cortical cells irrespective of whether it is within the fovea or outside [82]. Hubel and Wiesel, working on the visual system of macaque monkeys claimed [83,84] that the average receptive field image (and point image) diameter doesn't change with translation across the surface of the cortex and is approximately 2mm across.

These claims have been challenged in the intervening years [85,86] but recently Wassle *et al* [87] claimed to have resolved the controversy over the constant size of aggregate cortical point images. They were able to show that there are more than three ganglion cells per foveal cone in the primate retina. (The overall ratio is about 1 million ganglion cells for 120 million rods and about 6 million cones). They also showed that the ganglion cell density changed by a factor of 1,000-2,000 between the fovea and periphery. Because this result is within the range of estimates of cortical magnification factor there is no need to postulate selective magnification of the fovea either in the LGN or in the cortex. The distorted representation of the visual field on the cortex seems simply to be a manifestation of a simple rule which has been found to hold true for other sensory modalities: the central representation follows the peripheral receptor or neural density. This means that the architecture of the retina is alone responsible for the fact that the fovea accounts for such a substantial fraction of the cortical map. It also underlines a growing impression that cortical processing is initially relatively uniform, both within and between modalities – the structure apparent in mature organisms may be due almost entirely to the architecture of the associated sensory system and the nature of the information[7] impinging from the external world [66]. We shall return to this point again below.

Of course speculating about how and why the cortex develops in such a way says nothing about what the system is doing. The architecture of the retina has evolved so that one part of the visual field is sampled (by photoreceptors) and represented (by the ganglion cells) with a substantially larger number of units than the remaining visual field. Because of the nature of distributed representations, this fact alone is sufficient to account for the high degree of spatial resolution associated with the fovea. It is a property of distributed representations that the accuracy of encoding a feature is proportional to the number of units viewing the feature from slightly different perspectives [75, 88]. What the overall system is doing is "magnifying" a particular part of the visual field by providing more units to analyze that part of the field in the retina. The cortex is unaware of this magnification in the sense that it must be pre-structured in any way to deal with it. It simply receives projections from large numbers of ganglion cells (via one relay at the LGN). Two nearby cortical cells in a region representing the foveal visual field will belong to nearby points in the visual fields. Two nearby cortical cells in a region representing part of the periphery will belong to relatively widely separated points in the visual field. From this we would expect that the pairs of cells would carry very different information and be correlated

[7]Read: correlations within the signal.

in very different ways. These facts alone may be sufficient to set the two pairs of cells on very different development paths.

Before leaving the discussion on point images there are a couple of points worth noting. The first is that the aggregate point images are very large compared with the spatial grain of the observed cortical structure of hypercolumns and blobs etc. which is discussed below. There seems to be no fixed relationship between point images and the cortical architecture. Because of the substantially vertical (many-layered) structure of the cortex, the picture which the aggregate point image gives of cortical processing is too general and lacks precision. Just as problems arose with the "overwise" neuron because of an over-extension of the concept of receptive field, so too the notion of point image needs to be used with caution.

It seems that several different psychophysical phenomenon yield different equivalent cortical distances when scaled by the local magnification of the retinotopic map. This would seem to suggest that as well as considering as point images just mappings onto particular layers or particular populations of cells in the cortex it may also be useful to consider the point images of maps corresponding to particular visual (psychophysical) properties or phenomena.

3.6 Retinal Neurons – Processing, Function and Structure

3.6.1 An overview of the cat retina

Much of the work in the 1960s and '70s on the vertebrate retina was carried out on one species, the cat. As a result, by the late 1970s the overall architecture of the cat retina was believed to be well understood and results from it began to be used as the basis for computational models in computer vision [9]. This stage of development is reviewed here in the form of an overview of the basic features of the structure, processes and presumed function of the cat retina. It is not intended to be exhaustive, but simply to give a flavour for the level of detail and the type of processing concepts involved. At the time, it seemed as if the basic computational features would be common to most of the higher animals – certainly comparisons were made between these anatomical and physiological results from the cat and psychophysical results on human subjects. Somehow it seemed that the visual world was sufficiently in common (shared or mutual), sufficiently objective, that anatomical and physiological structures associated with vision might not have needed to be as adapted to a particular ecological niche as other anatomical structures such as limbs or mouth, etc. The notion of substantial commonality among different species is belied by the difficulty of making comparisons between, for example, the cat and monkey retinas on the basis of these early descriptions. The issues involved are discussed in section 3.6.2

The ability to adapt to light intensities varying over several orders of magnitude accounts for much of the functional architecture of the retina. From the rods and cones there are three separate paths leading to the ganglion cells in the vertebrate retina which mediate vision in daylight, twilight and starlight. The various properties of these paths are best characterised by the receptive fields of the bipolar cells which carry signals from the outer to the inner plexiform layer. In the dark adapted retina the receptive field consists simply of an *excitatory* centre and the bipolar cells act as spatial summators of single quantal events. In the light adapted retina the bipolar cells have a concentric *excitatory* centre and *inhibitory* surround structure allowing the bipolar cell activity to accurately represent local contrast. It is believed [89,90] that individual type *B* horizontal cells are responsible for the centre of the bipolar receptive field in the cat. The receptive field surround seems to be the result of the electronic interaction (via gap junctions), of many of the larger type *A* horizontal cells. A detailed description of the various known retinal pathways and the role played by the horizontal bipolar and amacrine cells can be found in Levine [11] and Hubel [61].

At the stage of the bipolar cells, the processing carried out by the light adapted retina seems to be restricted to a linear combination of local transduced intensities. There is one subtle difference between the centre and surround mechanisms which becomes important in the next layer of cells, the ganglions. The response of a particular type of bipolar cell, the CBb_1 cell, to light and dark bands in its receptive field *centre* are nearly equal, but of opposite polarity [89, 91]. The responses of this cell to light and dark bars in its receptive field *surround* are of the same polarity. It seems as if any change in a cone's output, whether an increase or decrease, is rectified to give an increased response in the bipolar's receptive field surround. What this type of detail illustrates is that in a very fundamental sense a *different type of seeing* is involved with the dark-adapted retina than the conventional interpretation of "seeing" which is in light-adapted conditions. The retina literally changes its processing structure to place a premium on detection rather than spatial shape and colour. In other words to interpret these changes in *computational theoretic* terms would necessitate a completely different computational theory for the dark-adapted case from the normal computational theory, and would also require a mechanism or description for the gradual movement between these two extremes. On the other hand the *information theoretic* approach invokes mechanisms and processes which are *invariant* throughout the transition between photopic and scotopic vision: viz. information transduction and transport and protection from noise (see section 3.7.1) and the light/dark adaptation can be seen to be a consequence of these rather than vice versa.

In the cat it is possible on the basis of cell morphology to distinguish between two different types of ganglion cell. The *alpha*-cell has a large soma and sparsely branched dendrites which form a wide field. The *beta*-cell at the same location in the retina, has a medium sized soma, and a smaller field of more densely branched dendrites.[8] The physiological distinction between the small receptive field and linear X-cell and the larger receptive field and non-linear Y-cell was mentioned above. There is now substantial evidence [89] to support the identification of the X and beta-cells and the Y and alpha-cells. These cells share the cone bipolar cells as a common input, but their substantially different connections to the cone bipolar cells gives the X/beta and the Y/alpha cells very different properties.

The X/beta cells collect about 50 synapses from each of about 4 cone bipolar cells with substantially overlapping fields, while the Y/alpha cells collect only a few synapses from each of up to 700 cone bipolars whose receptive fields overlap to a much lesser extent. These connection mechanisms between the cone bipolar and ganglion cells, seem to be sufficient to account for the very different properties of the ganglion cell types, despite their common input. In particular Sterling *et al* [89] argue that they account for the transient and non-linear nature of the Y/alpha cells, though this has yet to be verified.

It is interesting to note the extent to which the vertebrate retina has evolved to gain the most from the available dynamic range of neural processing, and to minimise the

---

[8] The cell size and dendrite field vary with eccentricity across the retina and it is therefore important to compare cells at the same eccentricity.

deleterious effects of noise from the quantal nature of transduction in starlight and from random thermal isomerizations of receptor molecules. Because noise is a particular problem in the dark adapted retina, the most significant efforts to offset it can be found in the rod pathways [89]. The problem of the limited dynamic range of neural processing and transport is common to all stages of light/dark adaptation and the retina uses two approaches to maximise the overall functionality. The one which is of particular interest to us is the extent to which the eye (and later the brain) exploits redundancy inherent in the incident visual patterns to reduce the amount or accuracy of data which needs to be processed or transported. From a theoretical point of view it is important to understand what mechanisms are used to reduce the redundancy of the visual data and how these mechanisms develop. If they are entirely genetically programmed – which is unlikely – we need to understand how they work. If they are not fully genetically specified then we must find out what it is, on top of the basic genetically derived structure, which allows the mechanisms to develop the way they do.

From a more practical point of view, it is instructive to see how the capabilities of the limited resources are maximized. Probably the most substantial and widespread adaptation is the separation of processing into an "ON" pathway and an "OFF" pathway. The "ON" pathway is one which involves, or responds positively to, a spatial or temporal increase in light intensity from the current adapted level. The "OFF" pathway is one which involves or responds to a decrease in light intensity. This dichotomization between something akin to "black" and "white" is also apparent in other sensory modalities as pairs like hot/cold, bitter/sweet, left/right etc. It probably arises because less metabolic energy is expanded if a nerve cell's normal firing rate is zero or low. Because there is no such thing as negative activity (in spiking cells anyway)[9], one population of cells will then be required to signal "positive" excursions above the adapted level and another to signal "negative" excursions below the adapted level. Note that the division between "ON" and "OFF" populations does not extend down to the transduction part of the photoreceptors themselves. It seems to arise at the

[9]The photoreceptors, horizontal cells, bipolar cells and most of the amacrine cells do not "spike" – i.e. produce action potentials. Information is conducted by means of graded or "slow" potentials whose physical values above or below the resting potential directly represent the required signal. In cells which produce action potentials (which can travel long distances), it is the frequency of the spiking which represents that signal.

receptor-bipolar synapse, possibly by different bipolars responding in opposite fashions to the same neurotransmitter released by the receptors [61]. The receptors themselves are unusual in the way they respond to incident light. In the dark the potential across the core membrane (maintained by molecular ion pumps in the membrane) is about -50 milliVolts, rather than the -70mV typical of nerve cells. When the core is illuminated, the potential difference increases and the receptor membrane hyperpolarizes, cutting down on the amount of neurotransmitter released. Stimulation seems to "turn off" receptors [61]. It is not known if there is a metabolic reason for this.

Another mechanism which allows for increased gain and greater dynamic range operates at the junction of core bipolars and X/beta cells in the inner-plexiform layer. "ON" and OFF" cells synapse in different parts of the inner-plexiform layer – ON cells in sublamina b and OFF cells in sublamina a. Two complementary types of core bipolar innervate the ON-beta cell in sublamina b – designated CBb1 and CBb2. The CBb1 cells depolarize and the CBb2 cells hyperpolarize to the same stimulus. The response of the ON-beta cell is caused by increased excitation from the CBb1 and decreased inhibition from the CBb2 cells in the manner of a "push-pull" amplifier [89].

3.6.2 More retinas, more functions, more complexity

In the 1970s it was believed that the five cell types contained in the retina – receptor, horizontal, bipolar, amacrine and ganglion cells – were relatively homogeneous, and completely defined the functional elements of the retina. By implication it was believed that understanding the interaction of these basic elements would completely specify the processing and function of the retina. It now appears that there may be between 50 and 60 distinct functional cell types in the retina [92] giving the lie to the notion of a retina doing little more than acting as a prelude for later edge detection. The amacrine cells alone have been categorized into up to 30 morphologically and functionally different types, many of whose function is almost completely unknown.

The concentric centre-surround antagonistic class of ganglion cell which Kuffler [72] discovered in the 1950s comprises a number of different subclasses characterized by ON/OFF, X/beta or Y/alpha, brisk or slow. However, all of these are still just one general type of ganglion cell. Another type is the directionally selective cell and there are many others with complex stimulus selectivities. Possibly the concept of a receptive field plotted by spots of light is not even an appropriate way of investigating the functioning of these cells. Ways of varying light intensity or pattern which more accurately reflect the natural degrees of freedom of their dynamics need to be found.

At the level of the bipolar cells the basic centre surround mechanism already exists. The transient nature of the Y/alpha cell response in the cat retina may be mostly attributable to the distributed nature of the bipolar cell input to the Y/alpha ganglion cell [89]. One role of some of the amacrine cells which are characterized by a very transient response may be to sharpen this response in the ganglion cells. This however is currently just speculation and still does not account for the large variety of amacrine cells [93]. It has been shown that amacrine cells with different shaped dendritic trees can be matched with an equally large variety (up to two dozen) of particular neurotransmitters existing in the retina – further emphasizing the fact that the different classes of amacrine cells must have different biological functions. Masland and his colleagues have described the properties of four of these cells [93,94,95] which give an indication of the diversity of functions that need to be carried out in the retina.

The amacrine cell with the clearest processing role of the four, is the cholinergic or acetylcholine-accumulating amacrine cell – also referred to as the "star-burst" cell because of its characteristic morphology. One of the most surprising facts about the cholinergic cell is the degree of overlap between neighbouring cells in the retina. In the peripheral retina, where the overlap is the greatest, a point on the retina is overlaid by the processes of up to 140 cholinergic amacrine cells [93]. It has been shown electrophysiologically that the cholinergic amacrine cell excites directionally selective ganglion cells, among others. Directionally selective cells respond selectively to spots of light much smaller than their receptive fields moving in certain directions within their receptive fields. However, the acuity with which these directionally selective cells can resolve small spots of light seems to be inconsistent with the size of the dendritic field of the associated amacrine cells. Masland suggests an explanation of this phenomenon, which if correct, could have far-reaching implications for the understanding of retinal processing mechanisms. He suggests that the dendrites of the cholinergic amacrine cell are locally electrically isolated and that excitation and response involve local branches of the dendrites, rather than the entire cell, its soma and processes, as is usually the case for neurons. This idea is supported by the length and thin-ness of the dendritic processes. The electrical activity of these cells is solely in the form of graded potentials which only travel quite small distances before fading away entirely. It is also supported by the fact that input and output synapses on these cells are typically situated side by side on the dendrites. Finally this way of functioning of the acetylcholine amacrine cells is consistent with what is known about the mechanism of directional selectivity [95] and provides an alternative explanation to those advanced by Koch, Poggio *et al* [96]. This issue has yet to be resolved, [97][10] but the possibility of autonomous local functions within regions of the dendritic tree of amacrine cells, i.e. organization at a subcellular level, increases enormously the degree and variety of function possible, as these cells mediate the pathway between bipolar and ganglion cells.

The second amacrine cell whose operation has been substantially worked out is the AII amacrine cell. This cell is characterized by having a narrow lateral spread of its processes and also by the fact that it appears to be bifunctional, operating with gap junctions and chemical transmitters in the ON and OFF pathways respectively. It is known to mediate the rod-bipolar to ganglion cell pathway for vision in dim light [89,90,92]. The indoelaminergic amacrine cell, of which there seems to be five distinct morphological types, is thought to have a major influence on the pathway by which dim light passes through the retina, though exact details of the mechanism remain to be elucidated [92].

The dopamine-accumulating or dopaminergic amacrine cells are notable by their scarcity and by their few long and thinly branched dendrites. The fact that the mosaic they form on the retina is sparse and full of holes suggests that they are not involved in any high resolution activities in the way that the cholinergic cells are. A diffuse overall control function is further hinted at by the fact that they only synapse with

---

[10]Recent findings, described by Miller [97] show that the cholinergic amacrine cells release neurotransmitters which are both excitatory and inhibitory. The directionally selective ganglion cells are known to receive strong excitatory cholinergic input. Pharmacological evidence suggests that the inhibitory neurotransmitter GABA also released by the cholinergic amacrine cells provides the inhibitory input in the null direction. It thus seems to be possible that a single cell type at a critical phase in the neural pathway may be responsible for both excitation in the preferred direction and inhibition in the null direction of directionally selective ganglion cells.

other amacrine cells. Dowling [95] has shown that the neurotransmitter dopamine can diminish the effectiveness of horizontal cells in mediating lateral inhibition effects. It does this by causing a decrease in conductance of gap junctions in the outer plexiform layer. One function of the dopaminergic amacrine cell may be to regulate the strength of lateral inhibition and thereby the centre-surround antagonism as a function of light/dark adaptive state. It may also mediate the overall excitability of inner retinal neurons. This indicates that in addition to there being three separate paths for vastly different ranges of light intensity, there is likely to be a continuous gradation of adaptation within each of these ranges. Instead of being stuck with a fixed centre-surround mechanism which must cope with large changes in light intensity, the centre surround mechanism may be continuously adapted to optimize its response depending on the ambient intensity at any time.

The picture that is now emerging is of a retina carrying out an amazing array of subtle and complex processes, with a delicately balanced system of controls which confer enormous adaptability. There are a number of possible reasons why the retina should be so complex [92]. One possibility is that the process of conversion of light patterns falling on the retina into an efficient and meaningful sequence of nerve impulses to be sent to the brain is extremely difficult. The factor which most compounds this difficulty is probably the need to operate with light levels varying over 10-11 orders of magnitude. Another possible reason for retinal complexity may be that what seem to be straightforward tasks may require sophisticated biochemical circuitry to implement in the retina. As long as we feel we understand what function a certain mechanism is carrying out there is nothing to stop us from engineering an implementation which is more suitable to the type of hardware and processing components available to us. It is important not to be too complacent about this approach however as the neural circuitry may be solving some problem of which we are not yet even aware.

The one factor which seems to distinguish the retina from other neural processing systems in the brain is the need to package a complex transduction and coding process into an extremely small volume. A number of mechanical and physical constraints force the retina to be very thin and force receptors to be as closely spaced as possible to minimize size and maximize acuity. Masland points out that if retinal neurons were the same size as those in the brain, the eye would need to be the size of a tangerine [92]. It is clear from the ability of the eye to operate at limits imposed by physical laws and from the enormous variety and complexity of its function, particularly in the inner retina, that the eye is a masterpiece of miniaturization and of evolutionary adaptation. The fact that it seems to be much more tightly specified by genetic factors and rigidly constrained in its development and architecture than the bulk of the brain is a hopeful sign. It suggests that we will be able to see to a much greater extent, the action of underlying developmental and information processing principles in the architecture and processing of the brain. In the case of the retina we seem to be constrained in the immediate future to discovering exactly what it does and why these functions are necessary and important. Epigenetic factors seem to play a smaller role than elsewhere in the brain.

3.6.3 The ganglion cells – a biological film?

The ganglion cells collate and transmit output from the retina. They form the layer of cells on the innermost surface of the retina (nearest the front of the eye), and their axons run across the surface of the retina, exiting as the optic nerve and travelling to the brain. Traditionally they have been the easiest to record activity from because of their location on the inner surface of the retina and because unlike all other retinal neurons (except possibly some amacrines) they "spike" or produce action potentials in their axons. These action potentials which code signal strength as pulse frequency are much easier to detect than the tiny changes in potential difference across the cell membrane which are involved in graded potentials

If any further evidence were needed to distinguish what a retina does, from the process of a camera, then it is provided by the ganglion cells. Instead of conceiving the eye as an optical device which projects onto a sensitive film from which neural "images" are transmitted to the brain, we are forced by the ganglion cells to modify our ideas. A more appropriate metaphor is one of the retina being made up of many "neural 'films' overlaid on one another, each transmitting a separate filtered version of the optical image formed by the eye" [98]. Now too, it is not "images" of intensity or colour which are being transmitted but things like motion primitives and contrast and data suitable for stereo and gestalt grouping on one hand or data suitable for detailed scrutiny on the other. We would again echo Barlow's sentiments:

> *... perhaps these arrangements should be taken as a hint of Nature that there is more to the problem of transmitting the image to the brain than can be understood from sampling theory. [16]*

In the cat, ganglion cells are normally classified as X,Y or W where the distinction between X and Y is made on the basis of their response to periodic gratings. The X and Y cells are believed to be most important in terms of pattern perception because of their high spatial sensitivity and the fact that their axons connect via the lateral geniculate nucleus (LGN) to the primary visual cortex. The term W-cell is basically a "catch-all" grouping which includes everything not classifiable as X or Y. The W-cells include several sub-classes, many of which have slow conduction velocities. One class of W-cells have a concentric centre-surround colour opponent response and project to the c-laminae of the cat LGN and to the superior colliculus. It is worth noting, because it has many properties in common with the cells in the colour-sensitive parvo system which dominates primate visual systems.

As well as distinguishing X and Y cells on the basis of their response to a periodic spatial grating there are many other factors which can be used as the basis for comparison. One in particular, which is related to this spatial grating response is their response to temporal modulation of a stimulus. The X ganglion cells response is modulated at the same rate as the temporal modulation of the stimulus. The Y ganglion cells however, respond with two components. One is a linear component which behaves like the X cell in responding at the same (fundamental) frequency as the stimulus modulation. (Incidentally, this component also displays the sinusoidal variation with spatial phase which is characteristic of the linear spatial response of the X cells). The Y ganglion cells also respond with a second, non-linear component at the second temporal harmonic frequency of the temporal stimulus modulation. This component shows no variation with spatial phase of a periodic grating and therefore is responsible for the lack of a null-position for grating reversal which is characteristic of Y-cells.

The behaviour of Y-cells is consistent with their response being due to the summation of many sub-units within their receptive field after some type of non-linear transduction process. These subunits look like the response of entire X-cells and probably correspond to individual core bipolar inputs. Sterling *et al* argue that the overall non-linear response of the Y-cells is a result of the lack of polarity of the core bipolar surround mechanism in conjunction with the large spread of bipolar inputs which the Y/alpha ganglion cell receives [89]. An alternative site for the underlying non-linearity is proposed to be at the bipolar-amacrine connection by Shapley and Perry [98] and they support this by examples of similar non-linearities in some lower vertebrates.

While both X and Y cells are sensitive to contrast, the Y cells are about a factor of three poorer in spatial resolution than neighbouring X cells. This poorer resolution is despite the ability of subunits of the Y ganglion cell receptive field to detect patterns with a similar spatial resolution as neighbouring X cells. It appears that Y cells signal the presence, and particularly the movement of small patterns within their receptive fields but because of the large spread of bipolar cells which are summed to give this response, the Y cells only locate the pattern imprecisely. Finally, the Y cells respond more transiently than X cells, particularly at high contrast and the average conduction velocity of Y cells axons tends to be slightly faster.

Although there is a large degree of similarity in the gross structure and function of vertebrate retina across different species, there are very significant differences which make identifications of particular classes of cells in different species difficult. This is the case with the X and Y ganglion cells of the cat and classes of ganglion cells in the primate retina labelled by the letters P and M (because they connect to parvo and magno cells in the primate LGN respectively). A firm identification of a correspondence, if any, would be very valuable because much work has been done on the visual system of the cat which could be carried over to the primate visual system if the similarities were sufficiently close. Briefly, the P-cells are the most numerous in the primate retina. They exhibit a sustained response to monochromatic light at the peak of the cell's spectral sensitivity curve but a generally transient response to broadband illumination. They have concentric centre-surround receptive fields and often show some type of colour opponency within or between the centre and surround. They send axons to four of the six layers of the LGN (2 for each eye). The M-cells also have concentric centre-surround receptive fields and give a transient response to broadband illumination. In contrast to the P-cells, they show little overt wavelength sensitivity though they may receive antagonistic signals from different cones. Their axons project to the two magnocellular layers of the LGN. In terms of spatial summation and spatial filtering, about 80% of M-cells in the Macaque monkey behave like X-cells with most of the remainder showing the non-linear behaviour typical of cat Y ganglion cells. Again there is a ragbag of assorted ganglion cell classes which are neither M- or P-type. None have been found to be selective to wavelength and they provide the bulk of the retinal input to the superior colliculus [98].

Principally on the basis of comparisons in terms of contrast gain, contrast sensitivity and variation of receptive field/dendritic field with eccentricity, Shapley and Perry have proposed a new correspondence between X, Y, P and M cells [98]. They suggest that the M ganglion cells and their magnocellular targets in the LGN are composed of two subpopulations, designated M_x and M_y which correspond most closely to the X and Y cells of the cat retina. The P group have no exact equivalent in the cat, certainly among the X and Y cells, but may be related to the colour sensitive ganglion cells grouped as W-cells. This replaces an earlier proposed equivalence between P and X on the one hand and M and Y on the other, which with recent evidence is seen to be untenable. Given this comparison it is very interesting to consider the functional properties of the various primate groups, M_x, M_y and P.[11] The behavioral contrast sensitivity of primates as a function of spatial frequency (see e.g. Levine [11]) and in its variation with retinal eccentricity has been shown to be consistent with the contrast gain and contrast sensitivity of primate M-cells. The high gain and high sensitivity of the M-cell pathway are probably important for pattern perception at low contrasts and at medium to low spatial frequencies. The high frequency tail of the behavioral contrast sensitivity may be due to the small P-cells which need high contrast to operate.

The P-cells have small receptive fields which vary little in size up to eccentricities of 5°, unlike all other retinal ganglion cells and unlike the scaling behaviour of contrast sensitivity or acuity functions. Small dense receptive fields are particularly good for achieving high acuity yet the P-cells do not have particularly good acuity due to their relatively small contrast gain. Shapley and Perry suggest that the P-cells may have a

[11]There may not in fact be two completely separate X-like and Y-like groups of M cells, but a continuum between these two extremes.

role in supporting the low gain, low resolution colour system and that the small receptive fields are for wavelength selectivity rather than acuity. If during development, cones with different spectral sensitivity are virtually indistinguishable, then the only way to guarantee that a ganglion cell achieves wavelength specific inputs, is if the receptor–bipolar–P-ganglion cells connections are all one-to-one contacts. This hypothesis is also consistent with the observed colour sensitivity and P-cell receptive field scaling with eccentricity. Beyond 5° the P-cell dendritic field gradually increases with eccentricity and many P-cells begin to lose the strong colour opponency which they exhibit inside 5°. Similarly, the ability to see a full range of colour is restricted to central vision and colour perception suffers substantially after about 10° out from the fovea. Shapley and Perry neatly summarize the situation:

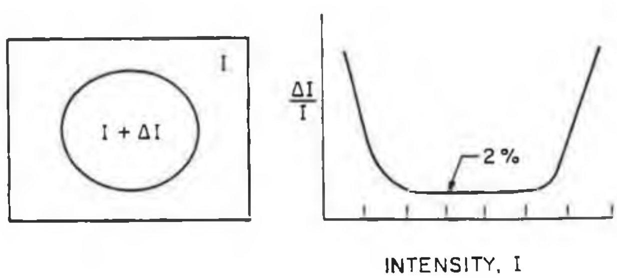
> *Cat X cells handle fine detail and are important for pattern detection while Y cells signal change and movement. Macaque monkey M cells report about fine detail and are important for pattern detection. Macaque P cells carry information about colour and about fine detail at high contrast. Trans-species comparisons may clarify or obscure these fundamental facts [98].*

3.7 Neural Coding

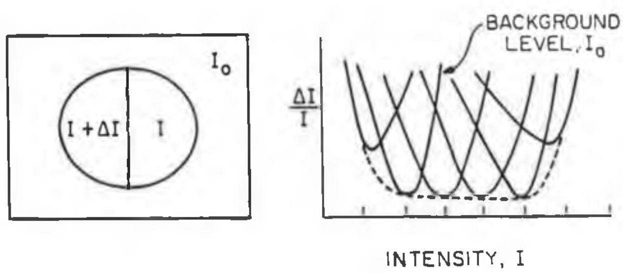
3.7.1 Noise in the compound retina of the fly

Surprisingly, it is work on the fly retina which may take us closest to understanding the overall purpose of the individual processing functions described above. The particularly modular structure of the fly's compound eye has allowed a detailed description of both the photoreceptor array and the neural wiring within the various components. Because the system is so well-defined, a careful analysis of form and function is feasible. Such a direct analysis is not currently possible in the retinae of higher animals because of the great number of unknowns still involved. Nevertheless, similar forms of processing seen in vertebrate retina do suggest similar overall function to that which has been described in the fly. It is expected that very similar design principles would operate in governing the transfer of information from photoreceptors to later processing structures – the principal differences arising from the type of information each species needs to extract from its environment and the biochemical activity it can support. In the case of the fly, the conclusion is that the principal function of the neural systems of the retina (which are analogous to the cones and bipolars of vertebrate retina) is information transduction and transport, and noise protection. To achieve these, a premium is placed on the efficient use of the available information, the available bandwidth and low-noise amplification.

Most natural objects are visible because they absorb and reflect constant proportions of the light falling on them in various parts of the visible spectrum. This is a physical property determined by the molecular makeup and surface characteristics of the matter comprising the object. It is a property which is invariant with normal changes in ambient illumination of natural scenes and is encoded by contrast or relative intensity (possibly as a function of frequency). Objects "look" the same to a contrast encoding system as the mean intensity changes over a wide range of values. The ability of individual photoreceptors to adapt their sensitivity to match the ambient light level is important for coping with the 10,000-fold range of light intensities during the day. This type of adaptation automatically allows them to encode contrast in their outputs



(a) No Background



(b) With Background

Figure 8. Schematic showing the range over which the Weber relation holds. Any of the curves for fixed I_0 in (b) is comparable with the dynamic range of most electronic imaging systems. From [60, p.33].

rather than absolute intensity. The ratio of just noticeable intensity difference to local mean intensity, called the Weber fraction ΔI/I [60], is found to have a nearly constant value of 2% over a wide range of intensities in human vision. As this ratio is equivalent to a small change in the logarithm of intensity, Δ(log I), then just perceivable contrasts at any mean intensity are directly related to changes in the log of intensity. Consider a response where inputs are considered equivalent if the quantity described by contrast, when superimposed on a background signal that increases with mean intensity, is constant. This type of response is equivalent to a logarithmic transformation of input intensity. It is a transformation which is found in a number of different types of insect photoreceptors and vertebrate cones [15].

Even though the photoreceptor response is matched to contrast which is an invariant of natural scenes, average contrast values are relatively low (0.4) and so the signal is still weak. In addition the contrast signal is superimposed on the relatively larger background signal which tends to use up the available dynamic range of photoreceptor response. The small fluctuations in membrane potential are therefore very sensitive to noise generated by neural processing and transmission. Two principal sources of noise have been identified in the fly retina: noise generated in the photoreceptors during transduction and noise generated at the synaptic transmission from photoreceptors to the large monopolar cells (LMCs).

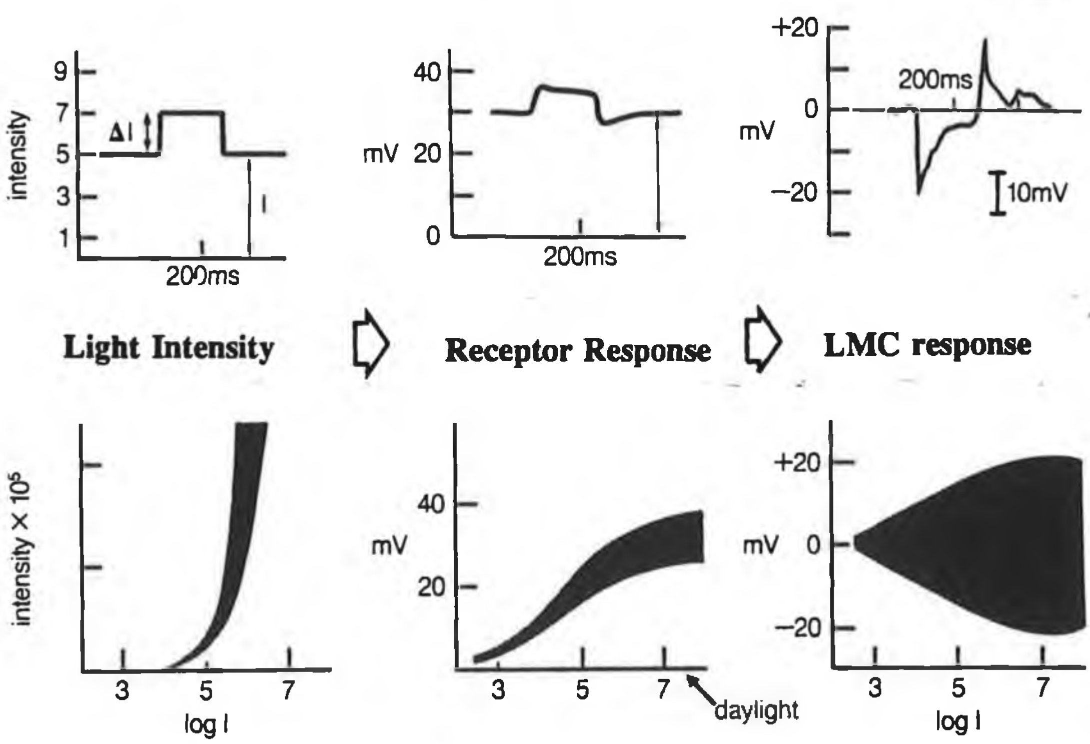


Figure 9. Diagram illustrating the waveform transformations and signal envelopes corresponding to most efficient use of the neural dynamic range. From [15].

These two sources of noise are about equal in the extent to which they limit signal quality during daylight. At any given intensity, summing intensity over a larger area to reduce photoreceptor noise will also have the effect of reducing acuity. The overall effects of synaptic noise can nonetheless be reduced by maximizing the amplification at the first synaptic stage after transduction which is the receptor-LMC synapse. Maximizing the signal-to-noise ratio by maximizing the amplification at this stage corresponds in physiological terms to maximizing the amount of transmitter released for a given change in receptor membrane potential difference,[12] but avoiding the saturation of post-synaptic response. Thus noise limitation considerations are sufficient to explain the widespread occurrence of high-gain synapses observed in sensory receptors [15]. Another device which both provides large gain and ameliorates synaptic noise (and is availed of in the fly retina) is a large array of chemical synapses between receptors and the succeeding neural processing stage (in this case LMCs). This array works on the principle that many synaptic connections made between a single pre-synaptic cell and a single post-synaptic cell transmit independently the same neural signal many times over. All of these signals are summed in the postsynaptic cell, providing a strong potential difference across the post-synaptic cell membrane simply at the metabolic expense of maintaining synapses.[13] As well as the summation over multiple synapses providing a large gain it also has the advantage that the quantization noise is introduced independently at each synapse into the identical copies of the signal being transmitted there and the averaging process reduces the effect of the noise.

3.7.2 Neural predictive coding

In addition to being amplified at the first set of synapses in such a way as to increase the signal-to-noise ratio, the visual signals undergo transformations in the retinal neurons of the fly which serve to increase immunity to noise by making more efficient use of the neurons' dynamic range. The signals recorded from the LMCs are made

[12]Since neuro-transmitters are released from synaptic vesicles in discrete amounts there is an intrinsic quantization "noise" inherent in synaptic transmission. If large amounts of transmitter are released for a given input signal, then this quantization effect is less noticeable.

[13]The pre-synaptic cell releases neurotransmitters which bind to receptor sites on the post-synaptic side of the synapse. These cause a release of energy in the macroscopic form of a graded potential – energy which was originally stored up by the metabolism of the post-synaptic cell. Simply at the expense of generating neurotransmitter molecules, the pre-synaptic cell can cause a large release of energy in the post-synaptic cell.

more transient by an antagonistic mechanism corresponding to the process of self-inhibition described in Limulus and similar effects in the vertebrate retinae. There is also an antagonistic response from neighbouring cells in the fly retina which are similar to the lateral inhibition which occurs in Limulus and the process mediated by the A-type horizontal cells in vertebrate retinae. The effect of these types of antagonism is to subtract away the background signal which remained after the logarithmic transduction process of the receptors and thereby allow subsequent amplification to expand the contrast signal alone to fill the dynamic response range of the LMC.

Laughlin [15] describes how these antagonistic mechanisms can be explained in terms of a predictive coding process which attempts to represent the visual data more efficiently thereby increasing both the capacity of the neural system as an information carrying channel and increasing the signal-to-noise ratio. Within any image of the natural world, spatial correlation will exist because, by the very nature of the external world, neighbouring points are more likely to have similar intensity values than more

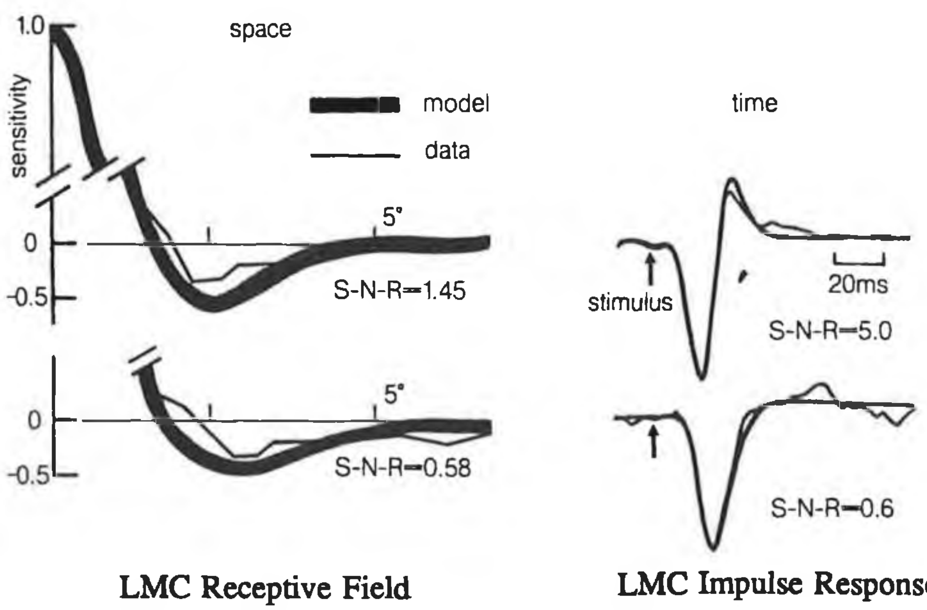


Figure 10. Illustration of the change in lateral- and self-inhibition corresponding to a change in signal-to-noise ratio and effectively different levels of light/dark adaptation. Adapted from [15].

widely separated points. There is also temporal correlation within signals transmitted by photoreceptors in an eye which is introduced by the relatively slow time course of these photoreceptors and also by virtue of continuity of existence in the external world. These correlations can be described statistically. Given this information and the intensity value of a particular element of a scene, predictions can be made about the possible values of neighbouring or succeeding scene elements. Predictive coding is the process of removing any components of a signal which can be predicted statistically from the context. The calculation of spatial and temporal statistical weighting functions for prediction within natural scenes [99] indicates two interesting results. Firstly, the weighting functions show little dependence on particular scenes. This finding runs counter to the intuitive notion that images of different scenes are so dissimilar that there could be little statistical relationship between them [100]. Secondly, there is a strong dependence of the statistical weighting functions on the signal-to-noise ratio in receptors, and therefore indirectly on light intensity. At the low intensities, randomness in the quantal receptor signal becomes more prominent and has the effect of decorrelating receptor signals. According to Laughlin, this means that the weighting functions must be extended to include a sample of neighbouring or preceding elements to give a reliable statistical estimate. By measuring the signal-to-noise ratio, impulse responses and receptive field at different intensities, it was possible to show [99] that predictive coding is a good model of the retinal process, so that for example, the extent of retinal antagonism changed with intensity as described by the predictive coding theory. The implication of this finding is that considerations of efficiency, noise immunity and matching to the statistical properties of natural scenes are close to the evolutionary pressures which conspired to bring the retina to its present state of development.

3.7.3 Matched gain

Returning to the problem of amplifying the visual signal produced by transduction in the photoreceptors: if there is too little amplification, the signal is lost in noise; if there is too much, then even ordinary low contrast inputs cause the output response to saturate. The ideal amplification process has a different gain for each intensity, depending on the frequency of occurrence – i.e. proportional to the signal probability. This type of amplification process helps to better distinguish intensity or stimulus values which occur most often. An alternative but entirely equivalent way of viewing the process is to consider the response range as being best utilized if all output values occur with equal probability regardless of the probability distribution of the input. In this way response values "contribute their fair share in carrying information" [99]. The desired amplification process is achieved when equal increments in response correspond to input changes equivalent to equal areas under the probability distribution curve of the input signal levels. This means that the input/output transfer curve, i.e. the relationship between stimulus contrast and LMC response in the fly, should follow the cumulative distribution curve of the input signal. This process of matching the amplification stage gain to the first order statistical characteristics of the input signal is exactly equivalent to the process of histogram equalization used in image processing as a means for improving contrast in a picture for human viewing. Considerations in terms of efficiency for information transport are rarely used in this context.

To summarize the material that has been covered so far: the retina of the fly has evolved to minimize the effect of synaptic noise. It achieves this by using predictive coding and matched gain to make more efficient use of the available neural dynamic range. These processes happen as early as possible in the retinal circuit at the interface between receptors and interneurons so that maximum effect is achieved. The predictive coding leads to a reduction in redundancy of the output signal from the retina. The antagonistic mechanism which implements the predictive coding seems to be able to adapt to the signal-to-noise ratio of the incoming signal.

Similar work, couched in exactly these terms has yet to be done on vertebrate retinae, including critical experiments which would indicate the relationship between lateral inhibition, receptive fields and the signal-to-noise ratio. The dopaminergic amacrine cells described above seem to behave in a manner which is in qualitative agreement with that required by the predictive coding model but the quantitative relationship which would link theory to reality has not been investigated. The bipolar cells in the vertebrate retina are the second order interneurons equivalent to the fly's LMCs and the ones which first exhibit lateral antagonism in their receptive fields. It is the response of these cells that we expect to see changing to match signal-to-noise ratio. The gross changes which occur between the light-adapted state for daylight and the dark-adapted state for twilight are too coarse to definitely support the predictive coding model. It would be much more interesting to see the effect of relatively small changes in ambient illumination completely within the light-adapted state.

The mechanism of colour opponency in human colour vision can also be explained using redundancy reduction arguments similar to those used for predictive coding. As described above the strong overlap in the spectral sensitivities of red and green cones allows them to be used as alternate elements in the sampling of luminosity. This strong overlap also introduces correlations between the red-sensitive and green-sensitive cone outputs which make coding in terms of parallel red and green colour channels very inefficient. Antagonistic colour opponent combinations have been observed in the concentric centre-surround receptive fields of primate ganglion cells. Behavioral studies of colour opponency also suggest that the human visual system codes colour in terms of spectrally opponent channels [101,102]. It seems that these mechanisms are compatible with the application of information theory to the reduction of redundancy in colour coding also [15, 101].

3.7.4 Alternative theories of neural function

Barlow's approach to understanding the visual system in terms of critical limiting factors is mentioned above in section 3.3.1 in the context of photoreceptor spacing. He proposes that the narrow dynamic range of neurons thought of as communication links, is another critical limitation which has a fundamental effect on the architecture and mechanisms of sensory information processing. As we have just seen, this approach in terms of communications theory, noise immunity and redundancy is successful in providing a model which corresponds closely with reality. It is a viewpoint that helps us to interpret the arrangements that we find in the visual system and hence improves our understanding of it [16]. The version of the approach presented in the previous section, so far claims little priority over other, possibly equivalent, explanations of the working of the visual system, like feature detection. It simply presents a model which is only supported in the extent to which it correlates with the reality of measurements on biological vision systems. There are however, theoretical reasons for a stronger stand on this issue. Laughlin sums up the situation in respect of the alternative and more traditional "explanation" of retinal processing in terms of feature enhancement:

> *... feature enhancement is often proposed as a major function of the retina. For example, retinal antagonism and amplification emphasize the timing of*

> *intensity changes and the location of edges. There need be no conflict between the principles of redundancy reduction and feature detection. In the absence of critical evidence, the coding efficiency argument embraces all feature enhancement arguments. Efficient coding will, by definition, emphasize every feature that carries information, including the location of edges and the timings of intensity change. To reject a coding efficiency argument in favour of feature enhancement one must not only invoke the qualitative benefits of enhancement, one must consider and evaluate the information that is lost or channelled elsewhere. [15]*

Barlow presents a somewhat stronger viewpoint on which of the approaches — the statistical redundancy one, or the feature enhancement one — has logical priority:

> *We often use contours to recognise objects, and lateral inhibition in the retina tends to promote the activity of ganglion cells at contours, where luminance changes rapidly. It is then suggested that the psychological importance of contours 'explains' the existence of lateral inhibition, but of course this puts the relation completely backwards: the brain can only use what information the retina supplies, so that the psychological importance of contours might* ***result*** *from lateral inhibition but could never* ***explain*** *it. Calling it an explanation distracts attention from the critical difficulties that must be overcome to extract useful knowledge from visual images, and it is understanding these limits that gives real insight into the organisation of the visual system. [16]*

Barlow is fairly unequivocal in his attitude to the "importance" of feature enhancement and subsequent detection as "explanations" of visual processing. Instead he stresses the need to understand visual processing in terms of reducing the redundancy of visual images by the "neat packaging of information" [16,103,104,105]. He suggests that the purpose of natural image processing is "to represent visual scenes by the activity of a sparse selection of reliable and non-redundant (i.e., independent) elements" [16]. Barlow's position is used here as a step to a much stronger view which puts the elimination of redundancy and the extraction of "invariants" in a pre-eminent position in the process of perception.

The whole point of an information processing approach to understanding sensory perception is that of all the processing configurations that the sensory system could adopt, it converges on one which is well matched to the statistics of its particular environment. As described by Torre & Poggio [106], visual perception is an ill-posed problem with no necessary solution. Biological organisms seem, as part of the strategy of overcoming the ambiguity inherent in "seeing", to have adopted the device of becoming "attuned" to the properties and structure of their surroundings. Thus we see in the fly that the ability of the retina to adapt to different signal-to-noise ratios and the matching of amplification to expected signal levels is a strategy which makes maximum use of the available neural signalling power. In a manner of speaking "assumptions" about the external world are built into the lowest level of neural processing. There is every reason to suspect that similar information processing principles are built into the processes of the cells of the vertebrate retina also. We find ourselves in close agreement with Laughlin in taking this idea a step further. Matching to the statistical properties of an organism's environment is not necessarily restricted to the lower levels of visual processing. In fact we suggest that it is a fundamental principle of all neural processing that neural subsystems in the brain are or become matched to the statistical properties of their input and indirectly to some properties or structures of the external world. Such a strategy may be interpreted in terms which are meaningful in our conception of our environment like edges, features, surfaces, volumes, objects etc., but these fundamentally do not explain the strategies. Again we take the opportunity to quote Barlow's aphorism: it is not "What is represented here? ... rather what types of information are brought together here?" [16]. There is no high level and low level vision with vastly different types of representation – there is only low level vision at different degrees of abstraction, i.e. dealing with information with different statistical properties.

The view proposed here however goes further than Laughlin. Inspired by Barlow's "activity of a sparse selection of reliable and non redundant ... elements" and by Watanabe's comparison of information processing and quantum mechanical systems,we believe that there is a much more fundamental reason for matching to the structure and properties of the environment than simply to get the most out of a neural transport mechanism (which it does). The elimination of redundancy at progressively more abstract levels, the detection of progressively more abstract invariants, the making explicit of implicit information *is* perception. There is nothing more, no homunculus in disguise "watching" what is going on. We return again to expand out these ideas below.

One very important question arises from this view of neural processing – what mechanisms determine the characteristics of neural processing circuits and how are they specified during development and maturation? For the visual system of the fly this question can be readily if incompletely answered. The development of fly lamina has been extensively documented and recent results suggest that both the performance and wiring of the visual system are influenced by early visual experience [15]. It is difficult to understand the position of the primate retina in this context as its development seems to be relatively complete by the time of birth. The situation for the primate cortex is more clear cut as effects of early experience on its development have been demonstrated.

3.7.5 Edges or artifacts: the Mach band phenomenon

The Mach band phenomenon, apparently the result of lateral inhibition in the retina, has inspired much of the work on edge-detection as a first step in visual perception. Barlow [16], Laughlin [15] and others have argued very strongly against the view that lateral inhibition subserves the task of emphasising contours or edges. Instead they claim that the Mach effect is simply a side-effect of the necessity to reduce as much as possible the gross spatial and temporal redundancy which exists in the visual data just after transduction. Laughlin describes how results on the SNR dependency of signal processing from the fly support the explanation of self-inhibitory and lateral-inhibitory mechanisms in terms of predictive coding for redundancy reduction. While the process of redundancy reduction is required to exploit to the maximum the information-carrying bandwidth of the rather noisy ganglion cells, it is not known if this type of processing is a pre-requisite for later stages of processing. Certainly contours are important in certain types of pattern perception but there is no convincing evidence that they are required for perception and not just a side-effect of coding for efficient data transport. Whichever is the case, it seems likely that the idea of a world consisting of a discrete number of self-identical objects is a view that is strongly influenced by the existence of the Mach band phenomenon [19, chap.2], and in turn by the peripheral coding which has more to do with the statistics of our environment than anything else.

3.8 Summary

Many useful ideas become apparent upon a careful analysis of the detail of research results that are available about the structure and operation of the retina. Thus we see, for example, that the form of the eye has more to do with the limitations caused by receptor cross-talk and organism size than striving to produce a faithful representation of the external world. It is very difficult to sustain notions about image-type representations when confronted with the details of the operation of the eye – particularly the eyes of invertebrates which nevertheless can have comparable acuity to ours. Also it seems that quite stable and quite powerful neural processing capabilities are relatively easy to generate, implying that there may be underlying structural dynamics or information theoretic principles which contribute to the final observed system. These principles or dynamics could be quite far removed from the apparent purpose of the system or unit. In other words, Marr's computational theory which emphasises the need to discover the role of a capacity: "what is being computed, and why", by ignoring either evolutionary or ontogenetic development, ignores a parallel system of structural modification – probably similar to what Varela refers to as a *metadynamics* [53]. This is related to the most important idea to emerge from the above discussion which is that much of the form and function of the retina can be understood in terms of information theoretic ideas. Because in turn the retina is more constrained (or "programmed") genetically than other parts of the nervous system, such information theoretic ideas may be even more relevant in these other parts, like the visual cortex.

We also see coming through, the important distinction between different functions of the visual system in terms of differentiation of into two or three different "strands" (magno/parvo initially) along the visual pathway. Much of the introspective idea of the way that the brain sees its environment seems to be strongly coloured by the properties of the scrutinising parvo system. A visual sub-system with these properties may be unique to primates. The magno system which is responsible for many of the less "solid" and more intriguing properties of vision like figure/ground separation, gestalt organisation, motion perception and so on, may be more representative of the visual capabilities of other vertebrates like the cat. This is yet another reason to try to erase from our ideas of how to build computer vision systems, notions of 2½-D sketches or 3-D representations.

Chapter 4

4 Information in Perception

(a)

> *Vision is the process of discovering from images what is present in the world, and where it is. Vision is therefore, first and foremost, an information-processing task, but we cannot think of it just as a process. For if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information – in all its profusion of color and form, beauty, motion and detail. [9, p.3]*

(b) In a book entitled *Visual Information Processing*, Spoehr and Lehmkuhle consider the human visual system as an information processing system – a system that can transform information from one form to another, that can reduce it (to avoid overload), that can elaborate it (filling in missing details), that can store it to memory and subsequently retrieve it [107, p.2]. Furthermore, they explicitly use a computer analogy to "refine our understanding of information processing".

(c) According to Marr [9, p.29], J.J. Gibson's important contribution was to note 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. He therefore asked the critically important question, 'How does one obtain constant perceptions in everyday life on the basis of continually changing sensations?'*

The answer according to Gibson was, in the direct detection of higher order variables or invariants:

> *These invariants correspond to permanent properties of the environment. They constitute, therefore, information about the permanent environment. [The] function of the brain, is not to decode signals, nor to interpret messages, nor to accept images, nor to organise the sensory input or to process the data, in modern terminology. It is to seek and extract information about the environment from the flowing array of ambient energy.*

(d) Consider the following which Dretske claims rests on the confusion of information with meaning:

*... something only **becomes** information when it is assigned a significance ... To speak of information as **out there** independent of its actual or potential use by some interpreter, and antedating the historical appearance of all intelligent life, is bad metaphysics. Information is an artifact, a way of describing the significance **for** some agent of intrinsically meaningless events. We **invest** stimuli with meaning, and apart from such investment, they are informationally barren.*

(e) Dretske [22, p.ix] puts the thing back into perspective:

*It is much easier to talk about information than to say what it is you are talking about. A surprising number of books, and this includes textbooks, have the word **information** in their title without bothering to include it in their index.*

4.1 Introduction

The extracts above are intended to give a flavour of the sense in which the term information is typically used in computer vision. If one begins to question exactly what it is that perception is, the vagueness and plurality of connotations of the term information becomes an immediate stumbling block. Information can have an objective and quantifiable quality as in the pulses on a wire, or the bytes stored in a computer's memory. Alternatively, it can have a more abstract, less quantifiable quality, as in the message or "news" carried by these pulses or in the bytes. And this ambiguity can be useful – it allows us to move from the concrete structure of the environment to the vagueness of meaning or ideas. The author can at each stage feel free to rely on whatever intuitive understanding the reader has of the term to convey their meaning. Like "knowledge" and "perception" it is one of these terms associated with (natural) intelligence which has so far substantially eluded definition or explanation in an objective context.

The aim of this chapter is to clarify some of the different quantitative interpretations of the term information as applied to signal processing, communications, the human visual system, pattern recognition and neural networks. Starting with the origins of the concept of entropy in physics, we first examine how this idea was generalized and adapted for use in describing discrete symbol systems in the work associated with Shannon. The key concepts described here are those of *surprise, equivocation* and *noise*, and we return to these several times in the remainder of this dissertation. A quite separate development of quantitative ideas about information was concerned with continuous or analogue systems and notions of sampling and quantization in these systems. This area provides the basis for the Gabor filtering and coding described in the next chapter and used there to interpret the filtering or coding properties of the visual cortex. The Shannon-type notions of information have also been used to quantify the human visual system (HVS) considered as a communication channel and this topic is discussed in section 4.4. A more general formulation of the physical concept of entropy than that used in information theory has been applied as an explanatory device in pattern recognition and pattern description and this is dealt with in section 4.5. Perhaps not surprisingly, the ideas of redundancy and equivocation have been shown to be very useful in describing the information processing function of certain types of simple artificial neural networks. However the simulation results achieved by Linsker, and described in section 4.6, present a great opportunity to begin to describe the developmental "forces" which are responsible for the information processing functions of natural sensory systems also. Finally having discussed many of the different aspects of quantitative and statistical interpretations of information in signals and in living systems we turn our attention to the issue of whether of not there is any other relevant aspect of the application of information to sensory perception, in a discussion of statistical and form redundancy in section 4.7. This topic is returned to in more detail in section 7.4, when we describe Dretske's semantic theory of information and relate it to the ideas being developed here.

The Collins dictionary defines information as

(i) knowledge acquired in any manner; facts; news or

(ii) as any data stored in a computer [108].

It is interesting to see these two definitions side by side. The first seems to imply that "knowledge" is internalised "information". That is, whatever the mind can acquire and store as "knowledge" – that is "information". The second definition – computer data – is often associated with the description of a computer as an "information processing system". The origin of this usage presumably arises from the loose identification of the terms "information" and "data". No doubt this is because of the usual assignment of meaning by people to the data stored in a computer. The juxtaposition of these two definitions for information above is no coincidence. Cognitive psychology, artificial intelligence and computer vision relies on, if not always the formal identification, then at least a strong association and some sort of interchangeability between these concepts [9,107,109,110,111]. This association itself is no coincidence – the modern concept of computation grew out of attempts to understand the brain as a logical machine [30,112]. It is unfortunate that the terms "information processing" and "computing" in the sense of the modern digital computer have become virtually synonymous. It is often useful to describe what biological sensory organs (including the brain) do, as the "processing of information", without implying that it is "cold", logical, unbiased functioning in the sense of a digital computer. Here we use the term "information processing" mostly in its more general sense.

There is a quantitative information theory (communications theory) but it literally tells us nothing about information – rather it deals with quantities of information, and not the information that comes in these quantities. Strictly speaking communications theory is not in its usage even directly concerned with amounts of information, but in how to characterize sources and channels for the flow of data. Still, it is possible that the ideas involved in the mathematical theory of communication could provide a foundation for relating semantic and syntactic aspects of what it is we call information.

A major driving force behind the study of human vision is the expectation that explanatory links can be forged between the experience of being able to see and the structure or processing of the visual system. In other words, while it is interesting in itself to explore and understand the workings of a complex system – and the visual system is certainly that – to find and understand neural correlates of subjective visual perception in a comprehensive way is one of the major outstanding unsolved problems [113]. One of the first major hurdles in this quest is the confusion that exists between the concept of information as it is used in science, the information that is represented by the activity and processing of neurons and neural systems and the information which the mind can conceive as knowledge and to which it can attach meaning.

4.1.1 Information in physics

The confusion does not stop with its interpretations outside of the so-called hard sciences. There are two quantitative uses of the term "information", one associated with physics and one, the one mentioned above, associated with mathematics (probability theory) and communications. The history and present usage of "information" in mathematics and physics is deeply intertwined with the concept of entropy introduced by Clausius [19, p.139] in the formulation of a conjecture – now called the *second law of thermodynamics*[1]. The ideas involved arose out of work on heat engines done by Carnot in the early 19th Century. In any system undergoing a reversible change, the change of entropy is defined as the energy absorbed, divided by the thermodynamic (absolute) temperature: $ds{=}dq/T$. The entropy of a system is thus a measure of the availability of its energy for performing useful work. This is a definition in terms of macroscopic physical quantities: work, temperature, heat engines and so on. With the advent of statistical mechanics as a microscopic level of explanation for thermodynamical quantities, the interpretation of the concept of the entropy of a system found a natural extension as a measure of the way in which the total energy of the system is distributed amongst its constituent atoms [114]. This re-interpretation of the thermodynamic quantity from an atomist point of view is due mainly to Boltzmann, Gibbs and Maxwell, working in the late 19th Century on statistical interpretations of mechanical systems [115]. In the statistical interpretation, the idea of entropy has been extended to include changes of the system that do not necessarily involve changes in energy. In general, the entropy of a system is used as a measure of its degree of "order" – the more disordered a system, the higher its entropy. In the subjectivist theory of entropy, originally due to Szilard [116], the entropy of a system increases whenever our information about it decreases. According to this theory "any gain of information or knowledge must be interpreted as a decrease in thermodynamic entropy: in accordance with the second law it must be somehow paid for by an, at least equal, increase in entropy". Entropy is *related to* the lack of information.[2]

[1]The second law states that entropy can only increase in a closed system. Experience teaches us to associate increasing entropy with the forward movement of time. Time "flows" in the direction of high probability, which is the direction of increasing entropy. The second law is statistical – "individual subatomic particles are conceived as such conceptually isolated short-lived entities that the second law does not apply to them ... Time reversibility exists *in potentia*, i.e., while the particles are represented by propagating wave functions [in Quantum dynamics]. Time irreversibility is an artifact of the measurement process [50, p.239].

[2]These ideas have recently been extended to logical computation, where a measure of randomness algorithmic complexity, is defined without recourse to probabilities (see Shannon's definition of entropy below). Algorithmic complexity sets limits on the thermodynamic cost of computations.

The first use of entropy outside of thermodynamics and statistical mechanics is attributed [19, p.139] to John von Neumann in his 1932 book on the foundations of quantum mechanics [48]. He introduced a quantity which he referred to as "microscopic entropy" defined as $S = -Trace(\sigma Log\sigma)$, to demonstrate the irreversibility of the process of physical observation. Here σ is a density matrix over an ensemble of identical quantum mechanical systems [115]. He did not make any reference in this treatment to order, or structuredness or co-operation. In a series of articles [117,118], Watanabe claims to have used von Neumann's "microscopic entropy" as a measure of structuredness or degree of co-operation between nuclear particles [19, p.140]. He used the term *"building block entropy"* to distinguish it from the usual thermodynamic entropy. It was shown to measure

(i) the degree of indeterminacy of the nuclear state of a single particle taken individually;

(ii) the degree of interaction and interdependence of particles constituting an organised system.

In the terminology of modern information theory (see below), the first measure corresponds to "ignorance" or uncertainty, and the second corresponds to "organisation" or "redundancy" [47, p.50].

In communications, Nyquist in 1924 [119] and Hartley in 1928 [120] had suggested that the logarithm of the number of alternative symbols (corresponding to maximum ignorance: $(ign\ E)_{max} = Log\ n$) could be used as a measure of communication. However, it was not until over 20 years later that Shannon and Weaver [121,122] put a mathematical theory of communications on a solid theoretical basis. As in thermodynamics and statistical mechanics, the term entropy again plays a key role but it should be emphasized that the uses are purely *analogical.* A further point worth stressing is that whether or not a message in a communication system is attributed a *meaning* by any external agency is irrelevant to its Shannon-information content:

> *The fundamental problem of communications is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently these messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to*

operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design[3]. *[121]*

4.2 The Mathematical Theory of Information

The mathematical theory of information, or communication theory or Shannon's information theory, should be carefully distinguished from any *semantic* theories of information. As mentioned above, the former terms are synonyms for a theory that deals with information in quantitative terms: with amounts of information and not the information that comes in these amounts [22, p.3]. As is articulated in the quote from Shannon, the task of information theory is to characterize message *sources* and data *channels*. Its methodology is to evaluate the amount of information identified with or generated by the occurrence of an event, and the reduction of uncertainty or the elimination of possibilities, represented by that event. Because the concept of Shannon-information is so important to the topics which are developed below, we will review some of the basic definitions and properties.[4]

A logical spectrum or partition [123] is a set E of propositions $\{E_1, E_2, \ldots, E_n\}$ such that any two distinct members are disjoint, i.e.,

$$E_i \cap E_j = \varnothing \quad \text{for } i \neq j,$$

and its members are exhaustive, i.e.,

$$E_1 \cap E_2 \cap \ldots \cap En = \square$$

We assume that at a certain *state of knowledge* we can assign probabilities $p\{E_i\} = p_i$ to each proposition or event E_i such that the probabilities p_i satisfy

$$p(E_i) \geq 0 \;\; \forall i; \qquad \Sigma_i \, p(E_i) = 1 \qquad (1)$$

We assume that the E_i's can be experimentally tested but that the probabilities p_i are assigned, before testing which one of the E_i's turns out to be true in this particular experimental test. Now, assuming that the test is made, our state of knowledge changes drastically. Now we know that one (and only one) of the E_i's has turned out to be true and we have a new "probability" distribution over the set of propositions or events:

$$p'_j = 1, \text{ for some } j; \qquad p'_i = 0, \; i \neq j \qquad (2)$$

---

[3]Dretske suggests that one source of the confusion which is endemic in the use of the term information is the fact that while between 1928 and 1948 American engineers and mathematicians began to talk about a theory of information in quantitative terms associated with communication, in the UK the usage moved away from communication towards a more general interpretation of the term.

[4]The terminology and notation follow those of Watanabe [47, p.8ff].

Considering again the situation before the test. The probability p_i is a measure of our expectation that E_i will turn out to be the case given our existing state of knowledge. If p_i is small for some particular i, we do not think it likely that the event E_i will occur. If in our test E_i does occur or the proposition becomes true, then our "surprise" is large. Alternatively, if $p_i = 1$ for some i then our state of knowledge guarantees us that E_i will happen in the test and we are not at all surprised when it does. Our "surprise" is zero. Any monotonic decreasing function $\varphi(p_i)$ can act as a measure of our surprise caused by the result E_i. Two useful functions are

$$\varphi(p_i) = -log\ p_i \tag{3}$$

and

$$\varphi(p_i) = -p_i \tag{4}$$

The logarithmic expression is more convenient and somewhat more intuitive because it satisfies an *additivity* constraint over two sets of mutually independent partitions [47, p.8ff]. If this property is not required then eqn. (4) is just as useful. The probability p_i means that we will suffer "surprise" $\varphi(p_i)$ if the event E_i occurs in the test. We expect that this event will occur with a probability p_i, however, and so the "expected (or average) surprise" is

$$E\{\varphi(p_i)\} = \Sigma_i\, p_i\, \varphi(p_i) = -\Sigma_i\, p_i\, log\ p_i \tag{5}$$

Given our "state of knowledge" before the test, then the expected surprise is also a measure of our *ignorance* with regard to the outcomes of the test:

$$ign(E) = -\Sigma_i\, p_i\, log\ p_i \tag{6}$$

The minimum ignorance occurs if we know that a particular event is certain to be the outcome of the test, i.e., if one of the p_i is unity (and all the others zero). The maximum value of ignorance occurs when we have absolutely no reason to think that any one event is more likely to occur than the other with our current state of knowledge. In this case all n events are equally likely with probability $p_i = 1/n,\ \forall i$.

Before the test, given our pre-test state of knowledge, we have some distribution of probability values p_i amongst the possible outcomes of the test. This state of knowledge can be assigned a measure of ignorance as in eqn. (5). After the test, the probabilities become the p'_i values of eqn. (2) with a corresponding zero measure of ignorance. The measure of information supplied by the test can be considered as the decrease in a corresponding measure of ignorance as a result of the test:

$$\begin{aligned} information &= decrease\ in\ ignorance \\ &= ign(E) - ign'(E) \\ &= -\textstyle\sum_i p_i \log p_i \end{aligned} \tag{7}$$

As Watanabe points out, even though ignorance and information have exactly opposite connotations, we end up with the same formula as a *measure* of each, in the case where the set of propositions E are testable. If the propositions E are not directly testable, the probability distribution of the p'_i's in eqn. (2) cannot be achieved in one test. Eqn. (6) however can still be interpreted as a good measure of our degree of ignorance or of indeterminability and (7) can still be interpreted as the amount of information or the decrease in ignorance supplied by any test that yields a new distribution over the p_i's.

The quantity $S = -\sum_i p_i \log p_i$ is usually referred to as a measure of *"entropy"*, by *analogy* with the thermodynamic entropy which increases with increasing disorder, chaos or ignorance. When the logarithm function is taken to the base 2, S is measured in units called "bits" where one bit corresponds to a pair of alternatives with equal probability.[5]

These ideas can be used specifically to describe what happens at *sources* of information which is one of the primary concerns of communications theory. If a source s has a number of discrete possible outcomes s_i with associated probabilities $p(s_i)$, then the average amount of information generated by this particular source, (the source entropy) is $I(s) = \sum p(s_i)\, I(s_i)$, where we now *define* the information obtained or generated by the occurrence of s_i as $I(s_i) = -\log p(s_i)$[6]. As well as sources, communications theory also considers data channels, generalising the treatment from the possible *outcomes* of individual events, to the quantitative evaluation of the possible *relationships* between events. Any event s with discrete, mutually exclusive

---

[5]In this context it is important to distinguish the *amount* of information (measured in bits) generated by a particular event, and the *number* of binary digits required to represent the possible outcomes of that event or state of affairs. In general, a redundant message will have more binary digits than bits of information, but in addition "bits" is a unit that can have fractional values while binary digits can only ever come in multiples of whole units.

[6]When discussing the notion of entropy itself we use the symbol S, but when talking about the information generated by an event we use the symbol I. They both refer to the same formula, just in different contexts or with different emphasis.

possible outcomes can be considered in isolation as a generator or source of information. However some situations or events r with discrete, mutually exclusive possible outcomes can be considered as *receivers* of information and more precisely as receivers of information about s. So instead of just being interested in the information generated by r, $I(r)$, we are concerned with how much of the information $I(r)$ is information received from, or about, the outcome of the event s. This quantity is labelled by $I_s(r)$. In other words, if there is a "reduction in possibilities" at r — what we have been calling an "outcome" — then $I_s(r)$ is a measure of how much of this reduction is related to the outcome of the event at s — a measure of the dependency between s and r. From this definition two important ideas arise:

Noise is defined as a measure of the information, or reduction in possibilities[7], at r that is *independent* of what happened at s. (It is always defined relative to some particular source [22, p.19]).

$$N(s_i) = -\sum_j p(r_j|s_i) \log p(r_j|s_i) \qquad N = \sum_i p(s_i)N(s_i)$$

Equivocation is defined as the information generated at s that is *not transmitted* to r.

$$E(r_i) = -\sum_j p(s_j|r_i) \log p(s_j|r_i) \qquad E = \sum_i p(r_i)E(r_i)$$

So

$$I_s(r) = I(r) - \textit{noise} \qquad \leq I(r)$$
$$I_s(r) = I(s) - \textit{equivocation} \qquad \leq I(s).$$

Now usually in communications theory the noise and equivocation are numerically equal in value, because $I(s)$ and $I(r)$ are defined to be the same: the *same* set of possibilities (or symbols) are used at both the source and receiver. In the more general case of interest to us here, this is not necessarily the case [22, p.239]. Increasing noise (the information available at the receiver which is independent of the information

[7]The "event" at r related to the arrival of information from s need not necessarily involve one single discrete outcome happening. It might simply involve a redistribution of the probability distribution over the possible outcomes r_i so that some become more likely and some less.

generated at the source), does not *necessarily* result in a reduction in the amount of information received at *r*. However, increased noise *can* eclipse part of the received signal, increasing the equivocation and reducing $I_s(r)$.

Dretske makes an important point that is often not immediately apparent in treatments of communications theory even though it is hinted at in the quote from Shannon above: because communications theory is concerned with sources rather than particular messages the primary quantities of interest are the *average* amounts of information generated by particular sources (the source entropy $I(s) = \sum p(s_i)\, I(s_i)$), rather than the amount of information generated by a *particular* event $I(s_i)$. Nevertheless in the context of epistemology it is precisely the amount of information generated by *particular* events that is relevant:

> *For when we talk about what can be known, we will be concerned, not with averages, but with the amount of information transmitted by (hence, the equivocation associated with) particular signals. [22, p.26]*

We return to this idea again in sections 4.7 and 7.4 below.

This treatment of the quantities of information associated with events, sources, and receivers, like most of Shannon's 1949 book [121], is based on discrete events. Shannon's communication channel models, accepted inputs and delivered outputs, at discrete instants in time, where the input and output symbols are drawn from a countable set. By means of coding theorems, he was able to rigorously establish the limits of information communication over such channels and these results could be expressed in numbers of bits transmitted per channel use [124]. The extension of these ideas to the continuous case of a model describing real communication systems was more troublesome. We examine this next.

4.3 Gabor's Theory of information

Shannon invoked the sampling theorem and used a "hand-waving" argument to claim in an imprecise way, that using signals of bandwidth *W*, one can transmit only *2WT* independent numbers in time *T* [124]. It was necessary to wait until the work of Landau and Pollack before the notion of *2WT* degrees of freedom for signals of

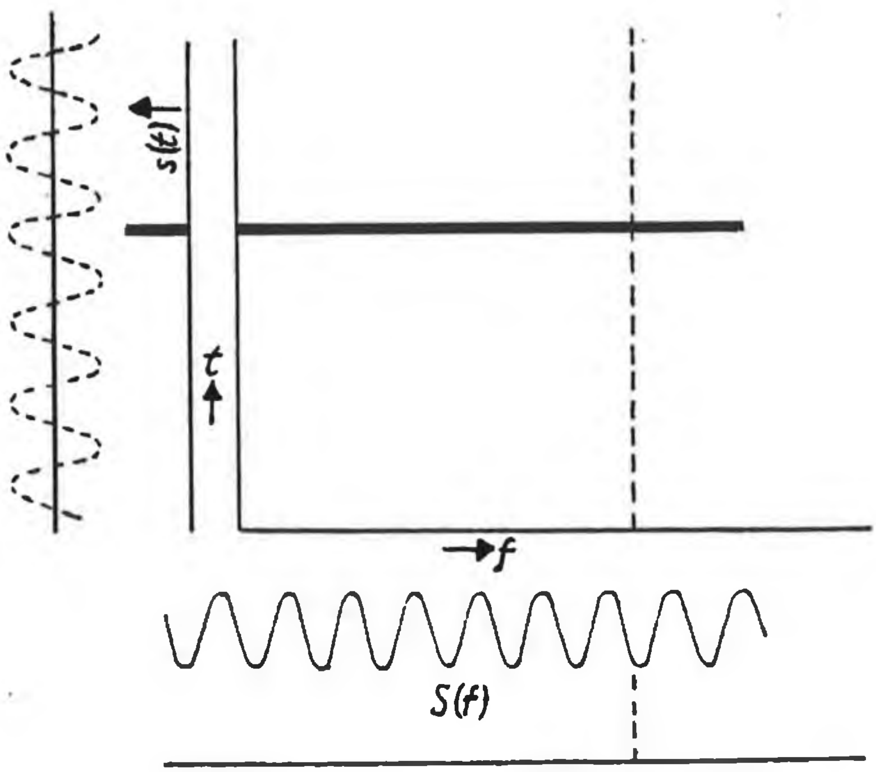


Figure 11. *Information diagram* as a Cartesian product of a time and frequency variable showing the representation of a pure sinusoid and a pure delta function. Adapted from [128].

duration T and bandwidth W was made rigorous[8] [125,126]. Nyquist [127] and Gabor [128] had earlier pointed out that approximately $2TW$ numbers were sufficient. Gabor's work is particularly interesting for many reasons. Firstly, his use by analogy of the uncertainty principle from quantum mechanics in signal processing indicates useful parallels between the two topics. His use of the term "information" is consistent with Shannon's later ideas but his method of application is very different. His introduction of time/frequency analysis paved the way for recent re-interpretations of the processing in the early visual system. Finally, he introduced an elementary function which accurately describes many of the characteristics of simple cells in the primate visual cortex (see chapter 5 below).

[8]David Slepian gives an excellent thought-provoking account of the paradox of real finite duration signals being band-limited in [124].

The motivation for Gabor's work published in 1946, and the related work of Ville [129] and Page [130], was a fundamental analysis and clarification of the physical and mathematical ideas needed to understand what a time-varying frequency spectrum is [131,132]. The application was the conveyance of information in communication channels and the optimal utilization of frequency bands. Up until that time communication theory was based on the two mutually exclusive idealized alternatives of time-domain analysis and frequency domain analysis.[9] Time-domain analysis employs operations at sharply defined instants in time. Fourier analysis employs infinite duration wave-trains of sharply defined frequency. Both of these methods of analysis are at variance with our intuitive notions of time and frequency, based on everyday experience – particularly our experience of auditory sensation like tones, and varying pitches in music, and also of human speech with its intricate modulations. The approach introduced by Gabor is to represent a signal in 2-dimensions with time and frequency as co-ordinates. This means having to find a 2-dimensional joint distribution, a function of time and a function of frequency, which describes the energy density or intensity of a signal *simultaneously* in time and frequency. Gabor called these 2-dimensional representations "*information* diagrams", as areas in them are proportional to the *number of independent data* which the "area" can convey. This sampling theoretic interpretation of time-frequency distributions was claimed by Gabor to arise from the fundamental "uncertainty relation" that exists between time and frequency description:

> *The frequency of a signal which is not of infinite duration can be defined only with a certain inaccuracy, which is inversely proportional to the duration, and vice versa. [128]*

Gabor defined "elementary signals" (now referred to as *Gabor functions*, though he called them *logons*) which occupy the smallest possible *area* in the information diagram[10]. These are Gaussian modulated sines/cosines, and any signal can be expanded in terms of them by a process of which time analysis and Fourier analysis

[9]Results using the short-time Fourier transform in the form of the "sound spectrogram" were published at almost the same time as Gabor's work. See [128, 129] for details. The problem with the short-time Fourier transform is choosing the time domain window-width suitable for a given time-varying signal.

[10]Because the Gabor *information* diagram is a Cartesian product of time and frequency, areas in the diagram are dimensionless quantities: pure numbers or scalars, literally data.

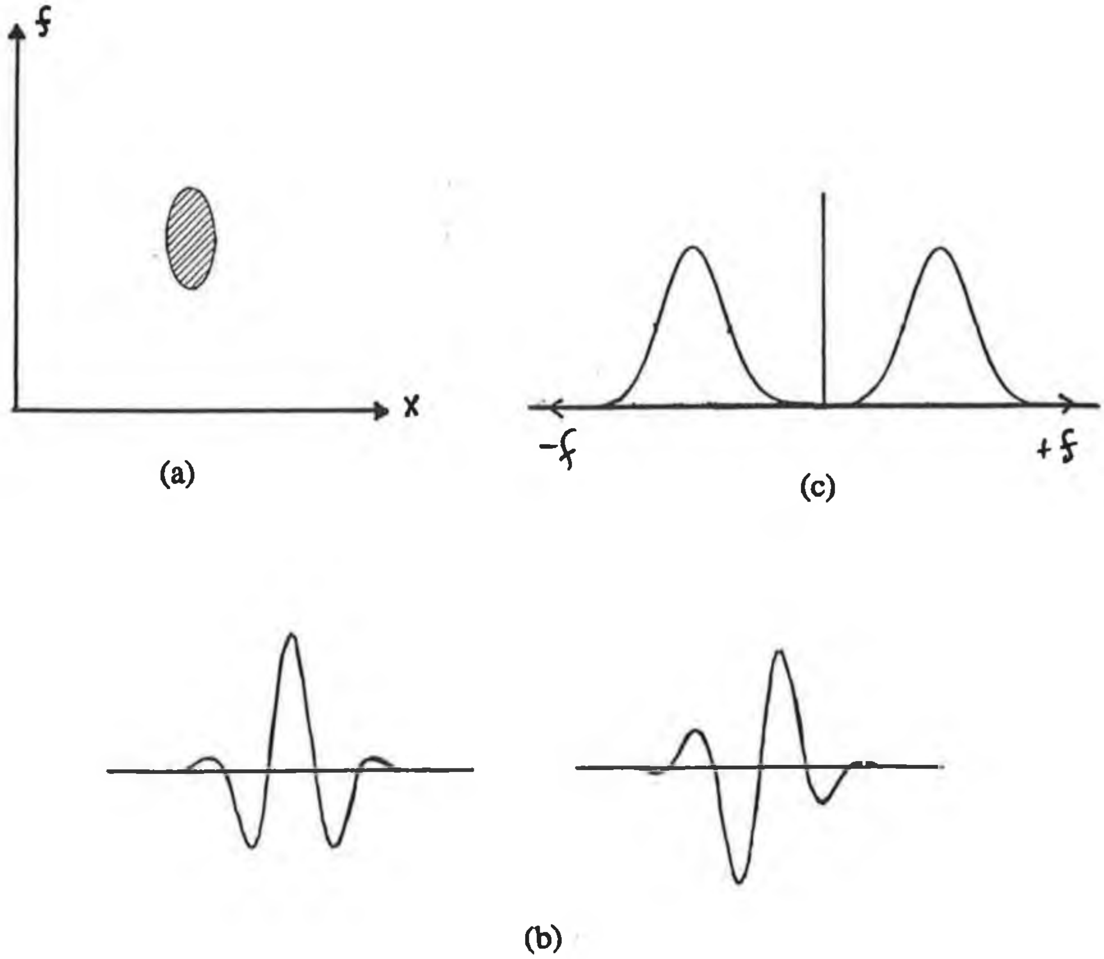


Figure 12. Illustration of Gabor elementary functions (a) on the information diagram, (b) in the time domain and (c) in the frequency domain.

are extreme cases[11]. Each elementary signal can be considered as capable of conveying exactly one *datum* of information.

At this stage Nyquist, Hartley and others had already concluded that the total amount of information that may be transmitted down a communications channel is proportional to the product of frequency bandwidth and time allowed for transmission. Gabor was attempting to put this result on a more rigorous footing, and at the same time give expression to more intuitive representations of signals. He described how no physical instrument for transducing sound can be represented either on the one extreme by a

[11]While the main advantage of the Gabor elementary function (GEF) scheme for representing signals is the achieving of the lowest bound on the joint entropy (defined as the product of effective spatial extent and frequency bandwidth), and the GEFs form a complete set, the GEFs are not independent. An analytical solution of the problem of finding expansion coefficients requires the introduction of an auxiliary elementary function as described in [132].

delta function, or on the other extreme by a pure harmonic oscillation. For every resonator ("oscillograph" or "reed"), finite damping times and tuning widths can be defined when the oscillations or response have fallen by some fixed amount (say *10dB*). Then there is a fixed relation of the form

Damping Time x Tuning Width = a number of order one,

and a characteristic rectangle in the information diagram with order unity area, for all types of resonators. Thus

> *physical instruments analyze the time-frequency diagram into rectangles which have shapes dependent on the nature of the instrument and areas of the order unity, but not less than one-half. The number of these rectangles in any region in the number of independent data which the instrument can obtain from the signal, i.e., proportional to the amount of information. This justifies calling the diagram from now on the 'diagram of information'". [128]*

Gabor attributes the fundamental limit, on the size of rectangles in the information diagram corresponding to physical instruments, to the making of a function of one variable – time *or* frequency – a function of two independent variables – time *and* frequency. He proves mathematically that this fundamental limit exists but offers no physical or intuitive explanations:

> *... as a result of this doubling of variables we have the strange feature that, although we can carry out the analysis with any degree of accuracy in the time direction or in the frequency direction, we cannot carry it out simultaneously in both beyond a certain limit.*

If Δt and Δf are uncertainties inherent in the definition of the epoch t and the frequency f, i.e. they are the damping time and tuning width respectively, then for all physical resonators we have

$$\Delta t \,.\, \Delta f \approx 1 \tag{8}$$

This expression bears a formal resemblance to the Heisenberg uncertainty principle of quantum mechanics. Gabor introduced the concept of the complex or analytical signal as the sum of the physical signal and its Hilbert transform in quadrature. He was able to use this function, along with the quantum mechanical wave-function formalism due to Schrodinger, Bohr, Pauli and others [46], and along with the usual procedure for deriving the Heisenberg uncertainty formula, to derive an uncertainty formula for signals described in time and frequency:[12]

$$\Delta t \,.\, \Delta f \geq \tfrac{1}{2}$$

where

$$\Delta t = \sqrt{2\pi\overline{(t - \bar{t})^2}}, \qquad \Delta f = \sqrt{2\pi\overline{(f - \bar{f})^2}}$$

are the "effective duration" and the "effective frequency width" respectively [128].

In quantum mechanics the minimum wavepacket is the wavefunction which simultaneously minimizes uncertainty in position and momentum representations. Since

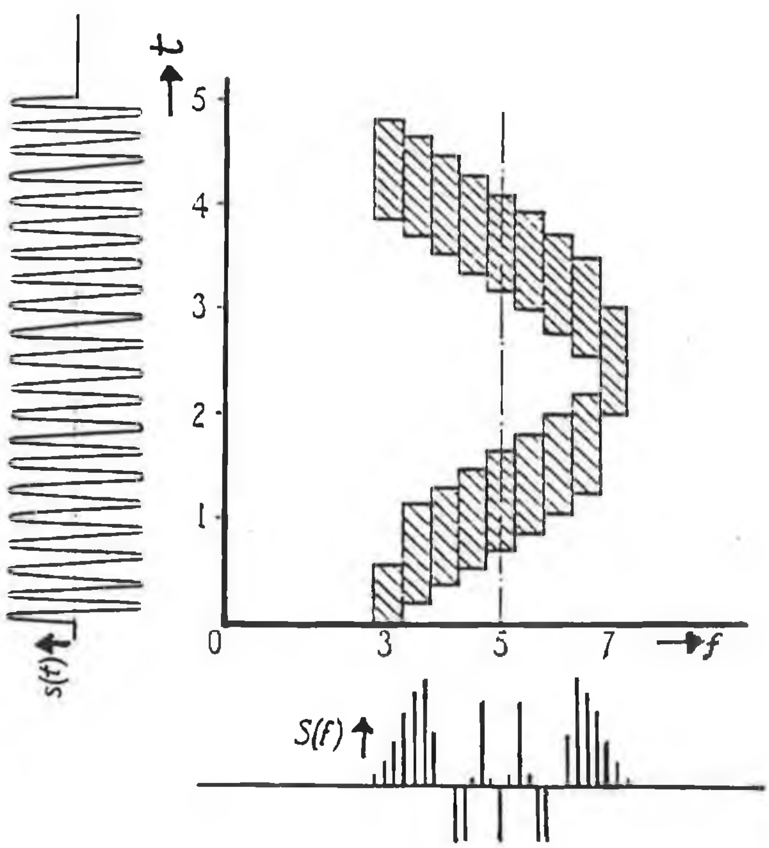


Figure 13. Illustration of a frequency modulated signal on the information diagram (centre), in the time domain (left) and in the frequency domain (below). From [128].

the mathematics of quantum kinematics and this time/frequency representation of

---

[12]Note that this result does not derive from quantum mechanics. It does not arise because of the quantal state of matter at microscopic levels. It is simply the result of a formal mathematical analogy which exists for underlying reasons that Gabor does not pretend to explain.

signals is formally identical, the same function describes a communications signal which is simultaneously localized to the greatest possible extent in both time and frequency domains. The time domain description of the minimum uncertainty function, the GEF or logon, is a Gaussian modulated sine or cosine. In the frequency domain it is two Gaussian functions on opposite sides of the origin. It turns out that the information diagram gives an intuitively very satisfying description of phenomena such as chirps signals, frequency modulation (FM) or time-division multiplexing [128].

Just as in the spectrogram, however, there is a free parameter in this GEF signal representation scheme. In this case it is the ratio of the sides of the GEF/logon rectangle in the information diagram, (and it is related to the notion of Q factor for an oscillator). We shall see below that the primate biological vision system may have come up with an appropriate way of selecting values for this parameter based on the statistics of images of natural scenes. The specific type of approach used in biological vision systems is not usually one availed of by engineers in general signal processing or communication problems because the signal ensembles in this case are assumed to be so general, but the freedom to vary this ratio can be useful here also.

Ville [129] applied a distribution originally used by Wigner [46,p.422ff; 133] to characterise the quantum mechanical duality between the position and momentum of a particle, to signal analysis. The Wigner-Ville distribution is

$$w(t,\omega) = \frac{1}{2\pi}\int s^*(t-\tau/2)e^{-j\tau\omega}s(t+\tau/2)d\tau$$

As mentioned above, Gabor in the case of one time dimension, and more recently Daugman [134] in the case of two spatial dimensions have shown that the class of Gabor filters achieve a minimum value of uncertainty as measured by the product of effective widths in the spatial and spectral domains respectively. That is, the uncertainty (as it relates to entropy) is measured separately along each dimension of the joint spatial/spectral representation [135]. Recent results published by Jacobson and Wechsler have thrown more light on the issues of resolution and uncertainty [136]. They claim that uncertainty should be measured in a joint Cartesian product domain rather than being the result of separate computations over two independent dimensions. They also show that the spectrogram, the difference of Gaussian (DOG) representation (see below) and the Gabor power representations are all smoothed versions of the Wigner-Ville distribution and as such cannot improve on the resolution achieved by this distribution.[13] The Wigner distribution is in turn a member of an infinite set of a more general class of distributions described by Cohen [137].[14]

Like Gabor, Cohen [131] makes it clear that the quantum analogy evoked by Gabor, and Ville is a formal analogy only. According to Cohen, the similarity arises because the probability distribution describing the likelihood of finding a particle at a certain position, is given by the absolute square of a wavefunction (which is the solution of a second order partial differential equation). The probability of finding that the particle has a particular momentum, is given by the absolute square of the Fourier transform of the wave-function. By associating a signal in communication theory with the wave-function, time with position and frequency with momentum, it is found that both systems have identical marginal conditions and are formally the same, even though the variables have different physical interpretations.

4.4 Sampling and Quantization

4.4.1 Isopreference curves

The discussion of information in the previous two sections has mostly been in terms of signals, discrete and continuous respectively. It is possible to extend these measures of information to investigate and quantify the information capacity of the human vision system (HVS), considered as a communication channel. These results, based on visual psychophysics are the subject of this section. Huang [138, 139] investigated the human ranking of different images and image classes as functions of the spatial resolution (the number of pixels per unit length) and accuracy of quantization (the number of grey value quantization levels). He plotted *isopreference* curves on 2-dimensional diagrams with axes representing the number of bits per pixel m the number of pixels along one size of the (square) images N, by linking points corresponding to images with equal ranking. The results are shown in Figure 14, for images with principally low frequency content (a), and for images with power

[13]Another advantage of the Wigner distribution over power representations is that it encodes phase implicity. Even though the distribution is strictly real valued the original signal can be recovered to within a sign.

[14]See [131] for a thorough review of the various time-frequency distributions which have been explored within this set, a description of their properties and some application.

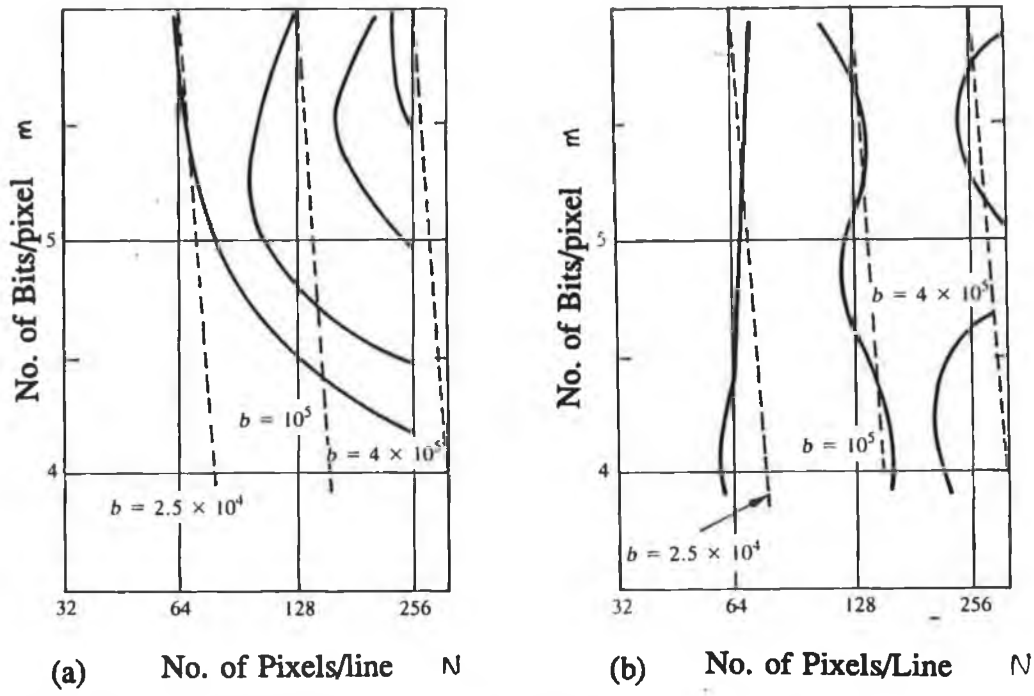


Figure 14. Isopreference curves for (a) most energy is in low frequency part of the spectrum and (b) energy is spread throughout the spectrum. Dashed lines are lines of constant bit rate. Adapted from Gonzalez & Wintz [139].

distributed throughout the available spectrum (b). Hardly surprisingly, he found that the quality of images tends to increase with *n* and *m*. This being said, there were a few instances, where for fixed *N*, perceived quality improved by ***decreasing m***, possibly because this increases the apparent contrast of an image. Secondly, he found that curves tend to be more vertical, as detail in the image increases (i.e. with an increase in power at higher frequencies). This result suggests that for images with a large amount of detail, only a few grey levels are required to represent the content of the image. Finally, Huang found that the isopreference curves depart markedly from the curves of constant numbers of bits $b = N x m$. Points on the isopreference curves can, with some care, be considered as representing images which, for humans, contain roughly equivalent amounts of perceptually (or semantically) relevant information. Lines of constant *b* represent what is, for the computer, constant amounts of data (also normally called information). The departure between the two again illustrates the difference between Shannon-information and semantic information.

4.4.2 The human spatial MTF

The spatial *modulation transfer function* (MTF) of the human visual system, has always featured strongly in psychophysical approaches to describing the capabilities and properties of the visual system [11, 64]. Its use assumes that at least to a first approximation, the visual system (including the optics, retina and the transduction and processing leading to a perceptual experience) can be modelled as a linear system. This assumption is not true [11,140,141,142,143,144], not least because of the logarithm-like non-linear compression which occurs in the photoreceptors. Notwithstanding this and other non-linearities, the first order approximation is sufficiently accurate in many cases for the MTF to be an extremely useful tool for quantifying aspects of the HVS.

For sinusoidal input, the spatial MTF is defined as the ratio of output contrast to input contrast as a function of spatial frequency. If it were possible to measure input and output contrast values directly, the human visual system (HVS) could be completely characterized (to this approximation) in the spatial Fourier domain. Unfortunately it is not possible to measure perceptual contrast output because of the many assumptions about the systems behaviour required to evaluate psycho-physical variables. Instead, the *contrast sensitivity* function (CSF) in the form of the threshold of contrast perception as a function of spatial modulation frequency $C_T(u)$, is used:

$$MTF \equiv H(u), \text{ assume } = CSF(u) = 1/C_T(u).$$

Roetling [145] showed how it is possible to use the MTF to make an estimate of the Shannon-information (in units of bits per pixel) that is retained by the visual system. He points out that normally in image processing for human re-viewing, the sample spacing and quantization levels are chosen so that the eye does not see degradations due to either process. On the basis of the high frequency cut off of the visual systems the sample spacing is assumed to be 20 pixel/mm or 60c/°. On the basis of the ability to perceive low-contrast differences in intensity at somewhat lower frequencies it is assumed that the visual system quantizes and represents signals to an accuracy of 8-bits per pixel. But this combined estimate is based on two different visual limits corresponding to two different points of the MTF curve, Figure 15. The estimate of the number of grey-levels required is made at a much lower frequency and therefore needs to be supported by a much smaller set of samples than the maximum sampling frequency. At the maximum sampling frequency, the visual system is very

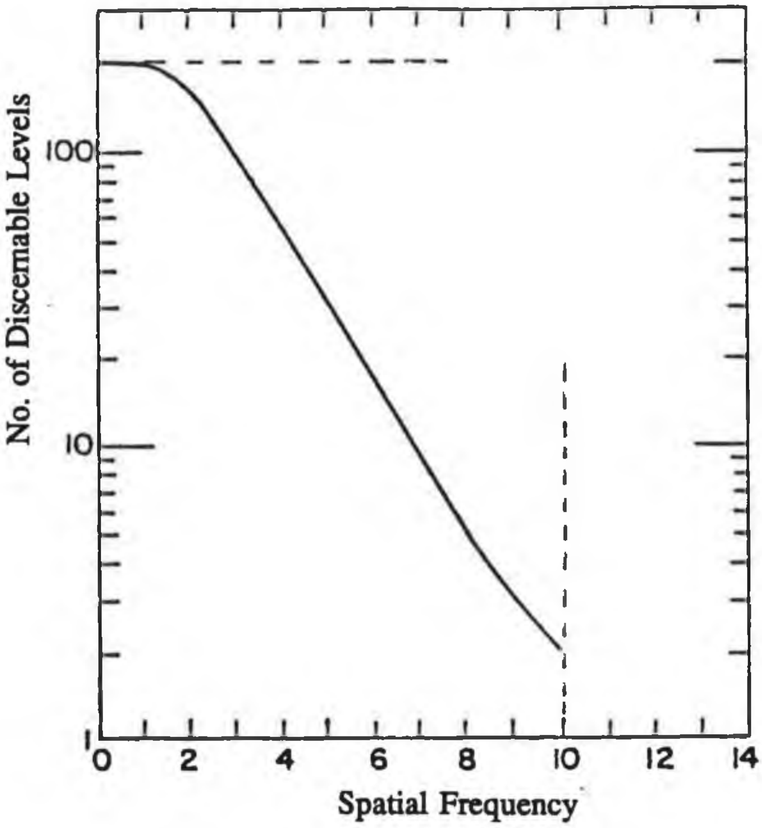


Figure 15. Visual contrast sensitivity (horiz. dashed line) and high-frequency cutoff (vert. dashed line) represented on the contrast sensitivity curve (with linear spatial frequency axis. Adapted from Roetling [145].

insensitive to contrast and only very few grey levels need to be coded. The combination of these limits, results in an over estimate of the "information" perceived by the eye.

Using a fitted curve to describe the MTF of psychophysical data, Roetling was able to show that an improved estimate of "visually useful bits of information" based on the entire MTF rather than just two points, is:

$$No.\ of\ bits/pixel \approx 2\pi\Delta^2(177.5)$$

$$\approx 2.8\ bits\ per\ pixel\ for\ a\ sample\ spacing\ of\ 20/mm$$

Briefly, he derived this result as follows. For the MTF he used the fitted curve [64, 146] which is normalized so that the peak of the curve corresponds to a just detectable modulation of 0.005:

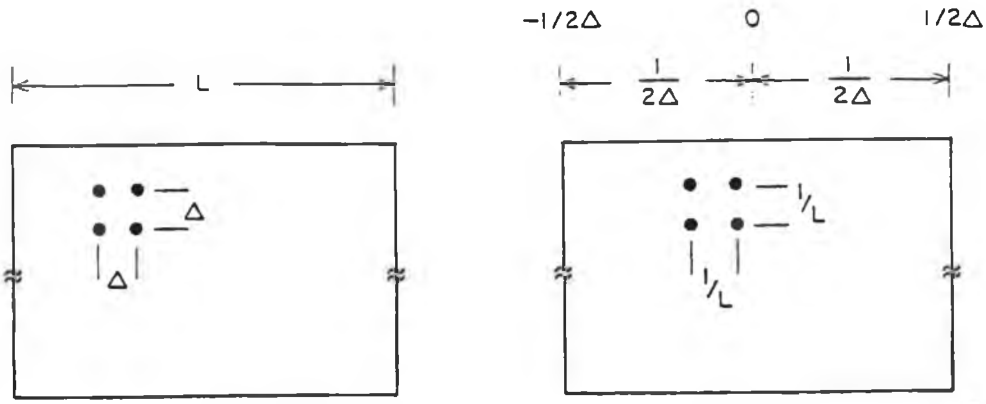
$$MTF = 5.05(e^{-0.138f})(1-e^{-0.1f}), \qquad f\ in\ c/^\circ\ .$$

Assume that at every spatial frequency, grey values should be quantized so that the just detectable modulation represents one quantization step. The modulation is

$$M = \frac{max - min}{max + min} = \frac{detectabledifference}{totalrange}$$

i.e. the number of intervals is the reciprocal of the just detectable modulation (original form of MTF data). So, (taking out normalization and converting to c/mm)

$$\#levels = 1010\ (e^{-0.138(5\upsilon)})(1-e^{-0.1(5\upsilon)}) + 1.$$



(a) Image Space

(b) Frequency Space

Figure 16. Sampling points in spatial and spatial-frequency domains.

The visual performance is described in the spatial frequency representation where an image with area $L \times L$ and sampling interval $\Delta \times \Delta$ is represented by a frequency range $\pm 1/2\Delta$ with frequency samples spaced at intervals $1/L$. (The total number of samples is the same in each case: $n^2 = (L/\Delta)^2$. The number of useful bits is the integral over the frequency range, of the product of the number of bits per frequency sample, $Log_2(\#levels)$, and the number of samples per unit frequency interval, L^2:

$$No.\ bits/pixel = \frac{1}{n^2}\int\int \log_2(\ levels)L^2 d\mu d\nu$$

The visual performance data has (assumed) circular symmetry so this can be rewritten as:

$$No.\ bits/pixel = 2\pi\Delta^2 \int_0^{\rho_{max}} \log_2(\ levels(\rho))\rho d\rho$$

where ρ is a radial spatial frequency. A numerical integration yields the required result.

Note that the number of bits/pixel $\propto \Delta^2$, so that assuming binary quantization of 1 bit per pixel, a sample spacing of 33.4 mm^{-1} would yield the same amount of visually useful information as the actual mechanisms of the eye. This is the reason why only

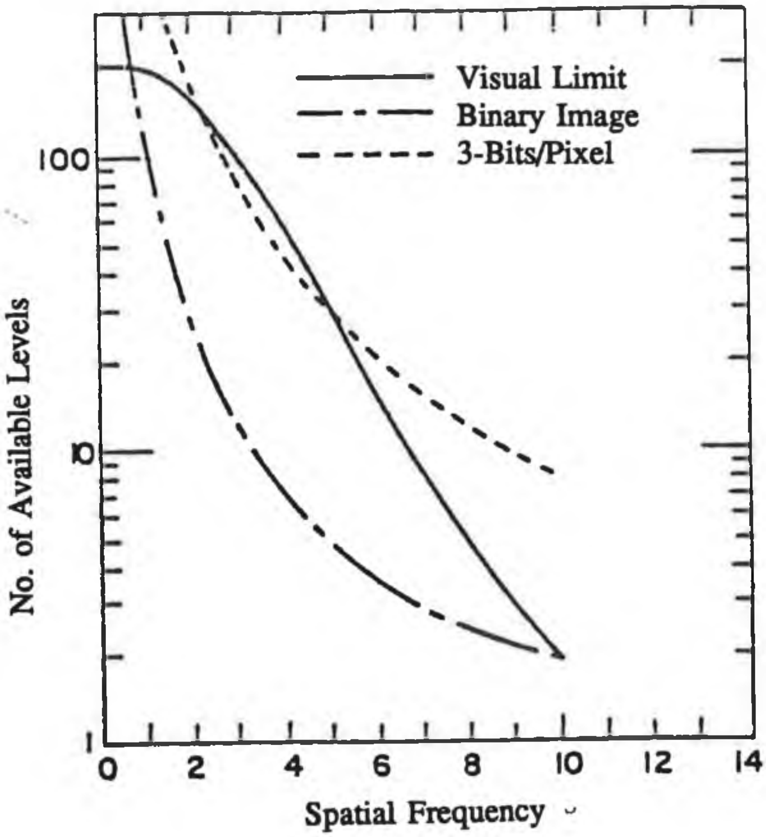


Figure 17. Low contrast reproducibility of 2 and 3 bit/pixel images as a function of spatial frequency with contrast sensitivity curve superimposed for comparison. From Roetling [145].

black and white dots in the form of dithered or half-tone patterns are able to reproduce full grey-scale representations if sampled at a sufficiently high frequency. Figure 17 shows the approximate equivalent low contrast reproducibility of different numbers of bits per pixel as a function of frequencies lower than the Nyguist limit. The human MTF curve is superimposed for comparison.

4.5 Entropy and Structure

The development of the concept of entropy from a purely thermodynamic quantity to the modern variants of the term is described above. The notion of entropy was subsequently used, by analogy, as one of the central planks in information theory. Its use in information theory does not however exhaust the possibilities of this notion as an explanatory device. An extension of the notion of entropy is useful for capturing the notion of the structure involved in pattern description and classification. This extension is based on Von Neumann's use of the quantity which he called *microscopic entropy*

$$S = Trace\ (\sigma\ Log\ \sigma), \qquad Trace\ \sigma = 1$$

where σ is a non-negative Hermitian matrix called the *density matrix*. (The density matrix is mathematically equivalent to the covariance matrix of probability theory). This interpretation of entropy (1932) is the first known use of entropic concepts outside thermodynamics. This expression is a more general concept than the usual definition of entropy in terms of the probability distribution of individual events only [123]. However, because the p_i are contained on the diagonal of the covariance matrix, the normal measure of entropy is implicitly contained in the microscopic entropy, which also in addition acts as a measure of second order redundancy. Von Neuman's use of the microscopic entropy to demonstrate the irreversibility of the process of physical observation was as a "simplified facsimile of the thermodynamic entropy" [47, p.518]. He did not use it in the context of structure, redundancy or co-operative phenomena discussed here.

Watanabe used the microscopic entropy of *each* nuclear particle in a system as an indication of the degree of their *co-operation* in nuclear matter. He called this use of the microscopic entropy a *"Bausteinentropie"* or building block entropy to distinguish it from the thermodynamic entropy. This Baustein entropy "plays the double role of a measure of [the] inexactitude of our knowledge of *each* Baustein (building block) and of a measure of strength of *mutual dependence* of constituent parts" [19, p.140]. This may give the somewhat confusing impression that on the one hand, the entropy of a constituent part contributes to the degree of organization while on the other hand structure is usually associated with low values of entropy, but, the existence of structure means that knowledge about a part allows us to more accurately predict the rest of the whole. "The possible variety of the state of the whole is restricted in spite of the [large] variety of states of the individual parts separately taken" [19, p.142]. Structure is maximized when the individual parts have large entropy but they act in consort so that the joint distribution has small entropy. The entropy of the whole is simply the sum of the entropy of the parts when there is no correlation or interdependence between the individual parts. If there is correlation or structure, this fact is measured by the disparity between the entropy of the whole and the sum of the partial or marginal entropies. The stronger the structure, the more the sum of partial entropies exceeds the entropy of the whole. Watanabe defines a function which captures a measure of the structure of a multipartite system:

$$J = \textit{strength of structure}$$
$$= \textit{sum of partial entropies - entropy of whole.}$$

If for example, the system consists of two parts represented by the variables x and y then

$$J = -\Sigma_x p(x) \, Log \, p(x) - \Sigma_y \, p(y) \, Log \, p(y) + \Sigma_x \Sigma_x \, p(x,y) \, Log \, p(x,y)$$

In non-physical applications, like image analysis, it often happens that the partial entropies do not appreciably change from one to another, so the structure often is a decreasing function of the entropy of the whole. Thus, the smaller the entropy (of the whole) the stronger the structure. In pattern recognition, this idea can be used as a heuristic principle: what Watanabe refers to as our "conceptual framework"[15] should be adjusted so as to maximize the structure, or in most cases, so as to minimize the entropy of the whole.

It is useful to explore this idea of structure further and to try to relate it to invariants and symbol systems in primitive perceptual observations. Consider the example of a pair of experiments; each experiment individually is a "part" and the joint experiment of the two taken together is the "whole". Independence between this pair of

[15]Recall that the process of making information explicit involves first determining a probability distribution over a symbol set which is equivalent to expanding our data vector in a certain co-ordinate system (axes are the eigenvector of the measurement operator corresponding to the symbol set). The choice between different symbols is easiest (the information or structure in the system is most apparent) when the probability distribution is as *uneven* as possible. This corresponds to *minimizing* the entropy of the whole as the symbols are defined in terms of the whole (or joint statistics), not in terms of the parts (or marginal statistics).

experiments means that there is no redundancy between them – our "surprise" at the result of the second experiment is not affected by the result of the first, and vice versa. Each time we carry out the second individual experiment we are getting new information about its outcome (our *a priori* uncertainty about the outcome goes to zero after the result becomes known). This is information which we had no reason to expect to know beforehand. The amount of information conveyed by the individual experiments is maximized and the structure function is zero.

Suppose now that there *is* some connection between the experiments – say we use one or other of two differently biased coins in the second experiment, depending on the result of the first experiment. Then, even before we carry out the second experiment we have an inkling of what the result is likely to be, and so are not terribly surprised on average when this result happens. On average then, we do not get as much new information as is possible in an unbiased experiment. There is structure in the experimental setup (witness our extended description). The results are redundant to some extent and so the J function is non-zero.

The combined or Cartesian product experiment is the largest common refinement of the two marginal partitions of the joint event space. The entropy of this joint, or product experiment – what Watanabe calls "the entropy of the whole" – is maximized if all the elementary events of the combined experiment are equally likely. This corresponds to the case of no structure and we should find that the entropy of the whole cancels out the sum of the partial entropies. On the other hand, if the joint probability distribution is as uneven as possible, the structure will be as large as possible, within the constraints of the partial entropies (which are based on the marginal distributions).

4.6 Linsker's Neural Information-processing Principle

Probably one of the most intriguing and original extensions of the use of Shannon-information to understand aspects of perception was work done by Ralph Linsker described in a series of articles published during the 1980s [18,147,148,149,150,151,152]. The basic problem he sets out to consider is the fact that unlike conventional computer hardware, neural circuitry is not hard-wired, but develops under the influence of genetic specification, and epigenetic factors such as neural electrical activity, both anti- and post-natally.

Much work has been done and continues to be carried out in embryology to investigate the mechanisms of the genetic specification of the development of the nervous system. In terms of epigenetic factors a number of specific details are known. Firstly, naturally occurring "noise", probably arising from thermal isomerisations in the photoreceptors, plays an important role in the *in utero* development of ganglion cell axons from the retina to the LGN in primates. Secondly, in the period immediately before and after birth, the visual cortex is being innervated by axons growing from the LGN, whose growth was in turn triggered by the arrival of the ganglion cell projections from the retina. There is a critical period during which the visual system must be subject to normal visual experience in order for the usual cortical structures to develop. Deprivation of visual experience in specific ways, such as blocking all light, or blocking all horizontal patterns to one eye, results in very specific corresponding deficits in the cortical structure [153]. These three factors, naturally occurring receptor "noise", early visual experience, and critical periods, are all known to be important in visual system development but the precise reasons why is what Linsker set out to investigate.

To motivate his approach he considers the following two questions [18]:

(i) What processing function does the neural "machinery" of the visual system perform on perceptual input, and what is the circuitry that implements these functions?

(ii) How does this "machinery" come to be?

On the basis that biological development processes are enormously complicated, and that anyway much of the detail is not yet known, his aim was not to model this development. Rather it was to see if there were underlying organisational or information-processing principles, which would explain some essential aspect of how the system developed or functioned, without necessarily having a neural model which was correct in all the details. If such principles could be discovered then they are likely to be generic to many different types of neural system carrying out diverse information-processing functions. Again this type of knowledge would be much more useful than a computational theory for the particular function because it would be less prejudiced by our particular values, or our particular observables, and closer to what is actually operational *for the system*, as opposed to our symbolic explanations. They might also give us an insight into exactly what quantities developmental rules or mechanisms act on, that are important to the information-processing function of a perceptual system.

The justification for the type of neural structure he investigates is as follows. The visual system of mammals is roughly organised into a multi-layered structure with connections within and between the layers. This is particularly true in the case of the early visual pathways including the retina, LGN and early visual cortex, The cells in each layer take their inputs signals from within that layer and from other layers and give a non-linear response which depends on these inputs and on the internal state of the cell. As well as the expected feed-forward of signals from the sensory surface towards "higher" layers within the nervous system, there are substantial *feedback* pathways where outputs from "higher" layers are used as inputs to the earlier stages. Nevertheless, if interesting self-organisational properties could be demonstrated in much simpler models involving feed-forward connections only, and involving cells

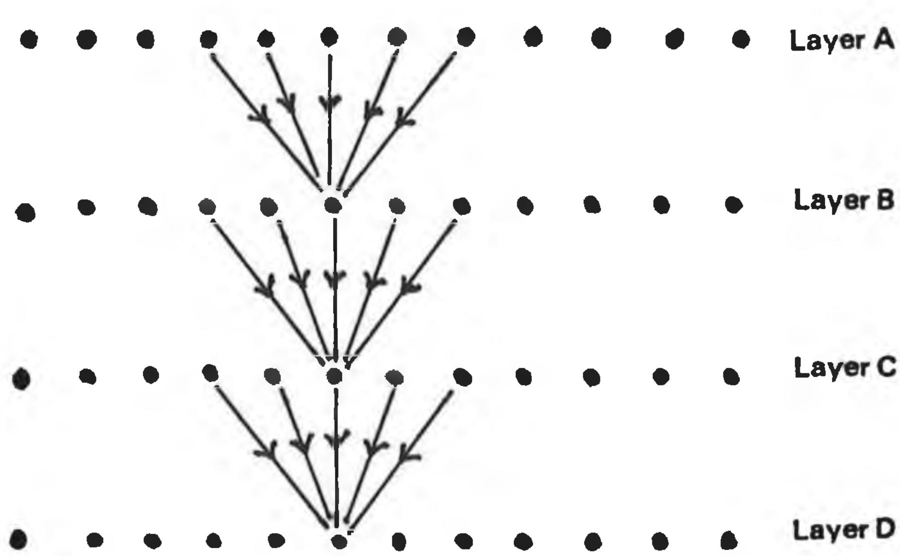


Figure 18. A schematic multi-layer feedforward network with linear nodes and random connections between layers.

with linear responses only, then we would know that these properties are not a function of the non-linearities or the feedback mechanisms. This knowledge is vitally important if we are to investigate what *is* the role of these latter mechanisms.

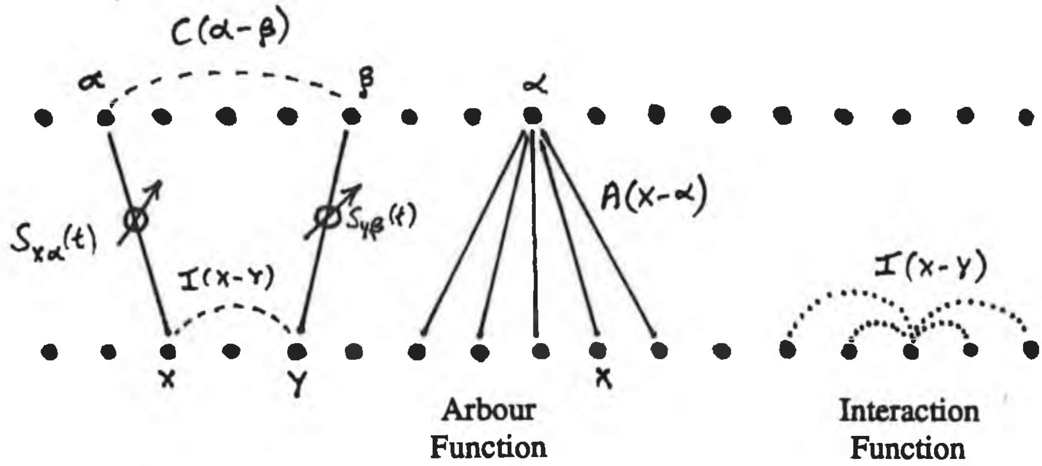


Figure 19. Schematic diagram illustrating the various symbols and functions used in the mathematical formulation of Linsker's network.

The simple multi-layer feedforward network[16] that Linsker uses is illustrated in Figure 18. Each layer is two-dimensional, with connections between the layers chosen randomly according to a distribution like a Gaussian. This ensures that most of the input comes from the cells in a neighbourhood immediately above the target cell, with few long range connections. The positions of these connections once selected, are not altered during an experiment, but the strength or weighting corresponding to each connection is adapted according to various rules like the Hebb rule and variants of it[17]. The input to the system as a whole is supplied to the first layer, (layer *A* in Figure 18) and could be actual visual input or random "noise".

[16]Note that any transformation implemented by a multi-layer feedforward network with linear nodes as described could just as easily be implemented by a single layer with the appropriate weights. But the point here is not to develop any *particular* transformation (cf. a computational theory), but to examine how the various layers of the network self-organize with random input or visual input to the first layer and using various weighting adaptation rules.

[17]The Hebb rule increases the strength of a connection between an input to a cell and that cell's output if the activity of the input is correlated with the output and vice versa.

A more precise version of the relationship between any two particular layers[18] is illustrated in Figure 19. Here α and β label cell positions in whatever layer is the input layer[19] for the connections being considered, while x and y label cell positions in the output layer for these connections. Each of α, β, x and y range over the full range of positions in the 2-D layers. Let $L_\alpha(t)$ be the activity of a cell in the input layer at α at time t and $M_\alpha(t)$ be the activity of a cell in the output layer at x at time t. The connections between the two layers are described by an "arbour" function $A(x\text{-}\alpha)$, and the connections between cells within the output layer are described by an "interaction" function $I(x\text{-}y)$. The input to position y at time t is given by

$$G_y(t) = \sum_\alpha S_{y\alpha}(t) f_2(L_\alpha(t))$$

This in turn affects the output of a neighbouring cell in the same layer at position x. Thus the activation at x at time t is

$$M_x(t) = \sum_y I_{xy} G_y(t) + constant$$

If $S_{x\alpha}(t)$ is the connection strength between a cell at position α in the input layer and a position x in the output layer then a typical Hebb modification is described by

$$\frac{d}{dt} S_{x\alpha} = \lambda A_{x\alpha}[M_x(t) - M_0]\, f_1(L_\alpha(t)) - \gamma S_{x\alpha}(t) - \epsilon A_{x\alpha}$$

Averaged over a large number of presentations this effectively becomes

$$\dot{S}_{x\alpha} = \lambda A_{x\alpha} \sum_y \sum_\beta I_{xy} C_{\alpha\beta} S_{y\beta}(t) - \gamma S_{x\alpha}(t) - \epsilon A_{x\alpha}$$

where $C_{\alpha\beta} = \langle f_1(L_\alpha(t)) f_2(L_\beta(t)) \rangle$ is the covariance matrix of the activities of cells in the input layer.

Now Linsker uses a Gaussian arbour function and ignores interaction within the output layer:

---

[18]The formulation described here is more general than that used by Linsker so as to accommodate the description of a system where the output layer can take input from two completely separate layers as in the ocular dominance model described by Miller *et al* [153], and also to describe the decorrelation ideas presented by Barlow *et al* [17].

[19]Note this is *not* the input to the system as a whole.

$$A(x-\alpha) = e^{-(x-\alpha)^2} \qquad I(x-y) = \delta(x-y)$$

and gets the following differential equation for the change of connection strengths over time (averaged over large numbers of presentations):

$$\dot{S}_{x\alpha} \propto \sum_{\beta} C_{\alpha\beta} S_{x\beta} + k_1 \sum_{\beta} S_{\alpha\beta} + k_2$$

What he discovered in simulations of this system is that (for certain combinations of parameters) when random noise is presented at layer A the cells at layer B adapt to calculate the local average of the activity in the overlying cells in layer A[20]. Once the cells in layer B have matured it is possible to fix their connections, feed their outputs to layer C and start adapting the connections between layers B and C. What is found is that cells mature in layer C, with what is effectively a *centre-surround* characteristic (or "receptive field") with respect to the original input in layer A. Continuing this adaptation through succeeding layers it is found that for certain parameter values the centre-surround characteristics become more pronounced, and at a particular stage cells develop which have *orientation selective* receptive fields with respect to the original input in layer A. Furthermore, if interactions are allowed between cells in a given layer the orientation selective cells become organised into patterns. This type of configuration is strongly reminiscent of the activity of cells in the layers of the early visual pathways in mammals, particularly the LGN and V1.

In a related series of experiments Miller *et al* [153] simulate the effect of having two separate sets of inputs corresponding to a *right* and a *left* view, with separate weighting functions $S^D = S^R + S^L$ and four separate correlation functions C^{RR}, C^{LL}, C^{LR} and C^{RL}. With Gaussian correlation within and between layers, (and also anti-correlation in certain cases), Gaussian and difference of Gaussian (DOG) layer interaction functions, and "box" arbour functions, they have demonstrated the type of "ocular" dominance patterns that are known to exist in the early visual cortex. Marshall [154] shows how Hebb-adaptive layered feed-forward networks with a natural propagation delay within layers, can develop quite sophisticated visual-motion processing capabilities (including direction and velocity sensitivity). He even suggests that it may be possible to show that higher-order capabilities, such as depth perception

---

[20]The connection strengths need to be limited so that they do not ramp to $\pm\infty$.

and object-recognition, can arise as self-orginizing properties in suitable network structures.

These results are tremendously interesting, although much work needs to be done to investigate and understand them fully. It could be argued that they are little more than structural mimicry, and do not even seem to carry out any particularly useful function to boot, what they demonstrate is that quite diverse aspects of the structure of a certain complex information-processing system (the nervous system) can be related to common and relatively simple adaptive mechanisms. This relationship puts a different complexion on attempts to explain these processing structures as separate aspects within a computational theory for something or other. It shows that there may be a more "primitive" level of explanation, nearer to the dynamics of development and activity of the information-processing system, where a *unified* account of the separate aspects can be given. The emphasis here is on *may*. Exactly what is happening in these simulations, and what this means, needs to be properly understood.

Linsker goes further than the simulations, and attempts to understand what the Hebb adaptations are actually achieving in information theoretic terms. Consider each presentation of inputs $L = (L_1,\ldots,L_N)$ as a message, with L_i denoting the activity of the i^{th} cell in the input layer, where the L_i values are quantized so that the N-dimensional space of L vectors is partitioned into boxes. (Two input messages are regarded as identical if they lie in the same box). If $p(L)$ is the probability that a randomly chosen message will lie in box L, the information obtained by selecting this message is $I(L) = log\ p(L)$, as described above. The average information conveyed is $-\Sigma_L p(L) log\ p(L)$. Now each input message L generates an output vector M which similarly lies in some box in an N-D space. Now the question is, if we know M, then how much more information do we need to reconstruct L? The answer is given by the amount of *equivocation* – the information about L that was not "transmitted" to M, i.e., the information we have about L given that we know M, $I_M(L) = -log\ p(L|M)$. The average rate R of transmission of information from the cell's inputs to its output is given by

$$R = < log\ p(L|M)\ /\ p(L) >$$
$$= <I(M)> - <I_L(M)>$$

So on the basis of an examination of the effect of the Hebb rules described above, particularly when the cell output is affected by noise relative to its input, Linsker claims that the network adapts to a transformation which *maximizes the rate of transmission of information from **L** to **M***, subject to constraints or additional cost terms. Furthermore he illustrates how depending on the level of noise, the *Infomax* principle organises a tradeoff between the benefit of having redundant ***M*** cell responses, which mitigate the information destroying effects of noise, or the informational value of having different cells evaluate a different linear combination of the input [18]. This description in information theoretic terms of the effect of a certain type of adaptation mechanism, yet again provides a different perspective on the "purpose" of perception: there is no need for any higher layer to attempt to reconstruct the raw sensory data or a "real" world representation, from what information it extracts. Rather the point is to enable the higher layers to use environmental information to discriminate the relative value of different actions. If the required information is discarded at any stage, it is no longer available for further use. Alternatively, if a local optimization principle is used at any stage, it cannot attempt to take account of global goals of the system and cannot know what information can be discarded. The principle of maximum information preservation ensures that the maximum information is transmitted through the system at all times, in a way which is neutral to the overall goals of the system. Remember the overall goals are not available at any level lower than the system as a whole anyway.

4.7 Statistical and Form Redundancy

One of the important features of communication theory in general, and Shannon's sampling and information analysis in particular, is that no account is taken of the semantic content of the messages. The information measures are based solely on choices between symbols, and probability distributions over sets of symbols. The quantities based on these information measures which are of concern, are averages over all the possible outcomes. There is a difference between the intuitive notion of information as the vehicle for knowledge, with all its connotations of meaning or semantics, and the measures of information in units of bits used in computer science and communications theory. The only type of information that we can quantitatively measure on the one hand is the Shannon information, while the only type of information associated with intelligence is the semantic variety. The tantalising similarities seem to only emphasize the gulf between the two. Bossomaier and Snyder [155] characterize the difference as a distinction between *statistical* and *form* redundancy. For example, in the English language, the letter q is always followed by the letter u; therefore in this context, the letter u carries no information (it does not help to discriminate symbols) and so can be omitted without loss. The letter e occurs more frequently in English than any other letter and so for example is assigned a short code in the Morse code. This is a statistical analysis of information. It involves correlations and frequencies of occurrence of letters or combinations of letters – first and higher order statistics – and it can be used to reduce *statistical redundancy*. It operates independently of words or meaning.

On the other hand Bossomaier and Snyder give the example of improving coding by using "a lexicon of English words". Coding can be made more efficient by rejecting nonsense combinations of letters which cannot occur (even though compatible with the statistics) or by introducing abbreviations which do not cause ambiguity. This is an analysis based on the semantic content of the message or text to be coded and it allows a reduction of *form redundancy*. It would seem that no achievable amount of logical computation can ever allow form redundancy to be directly analyzed on the basis of various orders of statistical redundancy. Yet there are reasons to believe that biological organisms are able to carry out this process. According to Bossomaier and Snyder, the processing of form information is a multi-level task which involves many operations in parallel. While very little other than this is known about the process, the authors claim that it is

> *greatly assisted by removing statistical redundancy and producing an economical representation at each level.*

They also argue that local spatial frequency analysis is the optimum strategy for removing statistical redundancy in vision.

One of the proposals considered here is that the removal of statistical redundancy is important, not because it assists later analysis, but because it is the first of many similar steps in the processing of visual data. The later steps (probably corresponding to the processing of form redundancy) have essentially the same nature as the earlier ones – like those that involve statistical redundancy reduction. The later stages have the same underlying nature, but are not identical with the early analysis for three reasons. Firstly, simply because the earlier processes have already happened and statistical redundancy has been removed, subsequent analysis will be operating on input with very different (statistical) properties from the input to the earlier processes. Secondly, the formation of cortical regions and the gross immature projections between them seem to be genetically specified and it seems to be this that determines the overall character of the sensory, motor and cognitive functions.[21] Thirdly the dynamics of signal flow within cortical regions, involving both afferent and reafferent travelling of information seems to play an extremely important, though as yet poorly understood role [37, 40][22]. One reason why the early processing stages of biological vision systems, seem to deal with the aspect of visual data that would normally be described by statistical measures, might simply be that at this early stage of processing, this is the only aspect of the overall visual data which is available to mechanisms whose connections are very restricted in spatial range.

4.8 Summary

There is always statistical redundancy in visual data in the form of correlations between nearby intensity values. These correlations arise from the finite size of objects in scenes and are increased by blurring effects such as diffraction and aberrations. Psychophysical evidence has shown [156, 157] that the human visual system utilizes at least 50%, and often much more of the statistical information in an image, regardless of form. The primary function of this chapter has been to begin to clarify the various quantitative usages of the term information, to apply them in precisely the context of a quantitative measure of visual information and to attempt to determine the extent to which they describe or quantify the information processing capabilities of biological vision systems. Starting with the notion of entropy, its development is traced through various extensions and re-interpretations, to the definitions of quantitative theories of information in the case of both discrete and continuous signals. The information theoretic notion of entropy is used in section 6.2 below as a heuristic

---

[21]It is interesting at this point to recall Crick's general description of the architecture of the brain in terms of a series of discrete maps which nonetheless interact to some extent at their edges. He also claims that when new functional areas of the brain arise in evolution, they arise in pairs [57].

[22]At the ESPRIT workshop in Killarney, Vision in Context, mentioned in the preface, the suggestion was made that "upwards" (or afferent) projection of information from the sensory surfaces, meeting the downward (or reafferent) flow of information from processing in "higher" regions of the cortex could be thought of using the metaphor of "controlled hallucination". Freeman's models seem to emphasise more of a dynamic "resonance" between these two flows of information [37, 40].

device and its expanded interpretation in terms of Watanabe's structure function helps to make clear the notion of statistical redundancy applied to pattern recognition. Linsker's work on simulations of simple neural networks leading to the *infomax principle*, allows an intriguing insight to the possible developmental mechanisms that cause the observed arrangements in biological vision systems to arise. These quantitative views of information are contrasted with semantic interpretations of information discussed by Bossomaier and Snyder and this topic is returned to again in much more detail when we discuss Dretske's semantic theory of information in chapter 7. Finally, the concepts introduced here allow us to begin to discuss different possible explanations or theories of perception, which is what we turn to next.

Chapter 5

5 Theories of Perception

5.1 Introduction

To properly describe the functioning of a biological visual system, one needs more than just a catalogue of different unit responses at different stages in the processing system. Not only do we need to know *how* neural mechanisms work at various physiological levels of abstraction, but also *why* that particular way is appropriate and *what* it is doing. Marr [9] was one of the first to clearly point out that description of the behaviour of neural system alone is not sufficient to allow man-made implementations – description alone in fact is dangerous: it subtly leads to mimicry. Explanation of behaviour in terms of the "whats?" and the "whys?" was the missing factor which accounted for much of the lack of progress in both biological and computer vision in the 1970s. His "computational theory" and "levels of analysis" is an attempt to fill the gap. Possibly because of the lack of time and the need to introduce a completely new approach to understanding natural and artificial sensory processing, Marr concentrated on explaining properties of the mature neural mechanisms, but in exactly the same way that we need an explanation of the mature mechanism's functioning as well as a description, so a description of how the mechanisms arise or develop is not the whole story. We also need to explain why they arise in this particular form carrying out this particular function. The computational mechanism which Marr describes as the first stage of his computational theory of vision calls for a very specific set of neuronal connections between different processing stages. To argue that such a specific programme of connections can arise, needs more justification than simply that the completed system is useful *a posteriori.* The work described here arose out of an attempt to allow an explanation of why the mechanisms for processing of visual information evolved and develop the way they must and apparently do. At least one important aspect of these problems of function and development can be illuminated by examining the issues in information theoretic terms. In the context of this chapter this means thinking of the visual data and the processing mechanisms in statistical terms. In chapter 7 we return to discuss a semantic theory of information which concentrates on the information in specific signals rather than ensemble or statistical averages.

Returning for a moment to the operation of the mechanisms for visual processing rather than their development, we need to construct a theory which explains this functioning. At this stage there have been several candidate theories some of which have been discredited for various reasons. The two interpretations of the properties of single cells in the visual pathway which have been around the longest are the rival theories of *feature detection* and *spatial frequency filtering* [56, chap. 3]. Historically these have had ample, if patchy influence on the development of computer vision. The concepts of, and basis for these theories are reviewed, in addition to arguments as to why they are untenable.

In many attempts to come to terms with the problems involved in perception, "invariants" have played a significant role. There are many different types of invariant. For example, some types of invariant can be described in mathematical or geometric terms, and computational ideas involving these have been gaining currency recently [158]. Another type of invariant can be associated with statistical constraints in signals or data, and can be detected or used by analyzing the data using concepts of redundancy or structure. The Gabor coding formalism for representing information, introduced in the previous chapter, can be understood in terms of the analysis or transformation of redundancy, and we deal with both of these subjects in some detail in the present chapter. Gabor filtering (which is closely related to the Wavelet theory of signal processing) has in fact, been used as a model for the response characteristics of cortical cells. The work of Field on Gabor-based codes, which is not intended to directly model cortical function but possibly to illustrate some of the information theoretic ideas involved, is described. The notion of an invariant is also closely linked with the idea of a symbol, and again we see a relationship between the Gabor ideas and the notion of symbols, uncertainty and multi-scale analysis described by Wilson and his colleagues, where they capture an information theoretic idea that does not exist in Shannon's framework. This is an issue that we return to in chapter 7. We do not pretend to fully grasp all the subtleties or implications of these issues but are confident that this is certainly a good place to look for solutions or explanations of the matters that are of concern here.

5.2 Feature Detector and Frequency Analysis Theories

The feature detection theory tries to interpret the input/output relationship of single cells as detectors of geometrical features involving a local analysis of the transduced visual pattern. The interpretation was originally motivated by the strong response which cells in the striate cortex of the cat show for edges or bars of luminance-contrast at specific orientations [159]. It was supposed that the sparse activity of "edge detectors" would produce an economical representation of the visual pattern. In later stages of the visual pathway it was found that cells seemed to become more specialized, needing more and more specific features to provoke a response, in what was described as a simple, complex, hyper-complex cell hierarchy [40, chap.5]. The location of the trigger feature needed to elicit a response also became less precise with progression along the pathway showing that later cells had larger receptive fields. These facts indicated the possible existence of a hierarchically organized visual system with increase in abstraction and decrease in localization with progression up through the hierarchy. At the base of the hierarchy were the precisely localized edge detectors. The top of the hierarchy it was expected to find cells which only responded when very particular objects – such as one's grandmother, or a yellow Volkswagen – came anywhere within the field of view.

With the increasing sophistication of knowledge about the visual system it has become clear that this interpretation of the response of single neurons in the cortex is untenable [160]. One of the first pieces of contradictory evidence was the demonstration by Stone in 1972 [161] that complex cells which were supposedly higher up in the hierarchy than simple cells because of their more complex response pattern and wider receptive field were actually driven directly by visual input from the LGN and responded *before* the simple cortical cells. A more crucial objection described in Bruce and Green [56, p.65] claims to refute the notion of feature detector altogether. To be a detector for a particular pattern a unit would have to respond to that and only that pattern at its input. In fact the cells typically give a range of responses which depend for example on the particular pattern at the input, its contrast or orientation. The rate of response only provides ambiguous information about the pattern of light in its receptive field, ie., it does not make explicit any specific type of information. Almost every part of a real visual image would have enough contrast to provoke some response in virtually all of the cells in earlier stages of the visual pathway.

One answer to this criticism is to reject the notion of single cells as feature detectors and claim only that they carry out some type of filtering operation. A further alternative position is to claim that single cells do not do anything which is identifiable by examining a single cell at all and that the real "movers" are groups of neurons acting in consort:

> *As members of local cooperative assemblies, they could collectively offer more exact descriptions ... More information may be represented by the pattern of temporal relationships between the firings of neighbouring units that by the firing patterns of any unit in isolation. [162].*

An even more radical position is to abandon attempts to explain the activity of cortical neurons in terms of a visual processing algorithm altogether and to rely on human ingenuity alone to devise a suitable computational theory of visual processing.[1] Despite the fact that a feature detecting theory is unsustainable as an explanation of the activity of single cortical cells in vertebrate visual systems, the principle of using edge detection as the first step in computer vision systems and theories is wide-spread in the computer vision community. The justification that edge detection has proved to be an effective means of coding many types of images is pragmatic, but somewhat of a "cop-out". There are many situations where edge detectors do not work, yet the human visual system does and we need to understand why, because these situations are as, if not more important than the ones where edge-detection does work. In addition an account compatible with biological vision systems might be very useful in the explication of later visual processing. A different account is quite likely to lead in the wrong direction.

A variation of the filtering position as an explanation of cortical cellular processing was inspired by the pioneering work of Campbell and Robson [163] using as stimuli, gratings consisting of sinusoidal modulation of luminosity contrast at different frequencies. Feature detection is a "local" process, involving only data from a very small part of the visual scene in each instance. Because of the global nature of the visual percept, some type of large-scale global analysis must take place sooner or later in the visual system. Then, the argument goes [56, p.68], if single cortical cells cannot be described as local feature detectors it may be because they do not carry information

[1]See for example, [56, p.86]. There seems to be an implication that whether or not Marr's theory is compatible with physiological findings, it is sufficiently well-grounded to stand on its own as an explanation of perceptual processing.

about local properties of visual pattern at all, but instead are involved in the processing of global properties such as *spatial frequencies*. (Its a weak argument!)

To investigate this possibility, physiological experiments were carried out to examine cortical cell response to spatial sinusoid patterns at the input, in terms, for example, of the frequency of maximum response, the selectivity to different frequencies or the sensitivity to contrast. Retinal ganglion cells and LGN cells were found to have broad spatial frequency tuning with little or no variation in tuning between cells with receptive fields at the same retinal eccentricity. This result correlated with physiological findings that receptive fields of these cells at the same eccentricity were usually very similar sized. In the cortex, things were very different. Both simple and complex cells were more narrowly tuned while each locus in the cortex contained cells with a wide range of optimal spatial frequencies. In the monkey, the optimal spatial frequencies ranged between 2c/° and 8c/° with bandwidths ranging from below one octave to about 3 octaves (with a median just greater than one octave).

The experiments of Campbell and Robson also inspired psychophysical investigations of the human response to spatial frequency gratings. This research culminated in the demonstration of the existence of multiple independent spatial frequency channels[2]in the human visual system and the proposal of a model by Wilson and Bergen in 1979 [164, 165].

While the physiological and psychophysical evidence of selective response to spatial frequency gratings is clear, the interpretation of the results in terms of a type of global Fourier analysis, like that of feature detection, cannot be sustained [71]. There are a number of reasons for this conclusion. Primary amongst them are the facts that the bandwidths are too wide and the responses too localized to support global Fourier analysis (see [56, p.71]). There simply is not the phase coherence across the image that would be needed to generate global Fourier coefficients, while phase sensitivity causes

[2]The original work suggested four spatial frequency channels. More have been described since, and even the suggestion of an unlimited number [135]. Recently, Barlow and Földiák have cast some doubt on the technique of response saturation which was often used in the investigation of spatial frequency properties of cortical systems [17]

ambiguity in single cell responses making the interpretation of their responses in terms of spectral component power very difficult.

Several suggestions (most of which did not amount to full theories of perception as the feature detection and Fourier analysis aspired to be) were made to explain the channel mechanism [56, p.72]. One of the most successful of the theories in terms of its explanatory power and subsequent exploitation is the idea proposed by Marr and Hildreth [59] that the multi-channel mechanism is an intrinsic part of an edge-detection process[3]. Marr objected to the two more traditional explanations on a number of grounds — the most important of which is the function which the theories *imposed* on the visual system in an attempt to justify properties of units within the system. Each of the two rival traditional theories — feature detection and Fourier analysis — is assumed to have as its goal the recognition of objects, not necessarily because of any *a priori* decision that this is the goal that needs achieving but because it is the logical conclusion of the assumptions made about single cells. In the case of the "feature detection" theory it is argued that pattern recognition is achieved by detecting the invariant set of geometric features that specify an object. In the Fourier analysis theory it is argued that recognition is achieved by detecting the invariant spatial frequency components that specify an object.

The problem with both of these theories is the "large degree of commitment of the visual cortex to a particular kind of abstraction of information; that required to identify objects" [56, p.84], when in fact there is a large amount of other information available from the environment about such things as the positions of objects, their depth and relative movement etc. In both of the traditional theories, this type of information is thrown away at an early stage in the visual pathway.

Marr claimed that his "raw primal sketch" still retains all the information about position and movement, etc., present in the input images. This means that it can in theory be used as the input to a wide variety of processes for recognizing objects, analyzing three-dimensional structure and computing motion. But, a careful analysis of the type of processing that must be involved in visual perception

---

[3]Field [71] presents an alternative explanation in terms of the statistical redundancy of natural images which we shall discuss below.

[166,167,168,169] and constraints on the computational resources [170] indicate that there must be form/motion subdivision of visual processing which is incompatible with Marr's generalized feature detection (edge; blob; orientated line segment, etc.) theory. Wilson *et al* show that attempts to simultaneously measure spatial position and spatial frequency are incompatible – measurements of position destroy all spatial frequency information and vice versa. The analysis of form requires the extraction of detail at high spatial resolution which implies a large spatial frequency bandwidth; (a delta function has infinite bandwidth). Measurements of velocity then result in large uncertainties due to the consequentially wide temporal frequency bandwidth. (Velocity is related to the ratio of temporal to spatial frequencies [170,171]). This reasoning is strongly supported by a growing body of evidence which describes the existence of a form/motion subdivision of the visual pathway in primates [13,172]. It is now becoming clear that the cortex, including the primary visual cortex V1, has a more complex set of interconnection, including multiple reciprocal pathways, than could possibly be consistent with any of the step by step theories of early visual perception mentioned – including Marr's $\nabla^2 G$ based raw primal sketch and full primal sketch models [173,174][4]. To an extent, the difficulty of building a complete theory of perception and the virtual impossibility of reconciling it with the growing body of knowledge of biological sensory systems, particularly vision, has encouraged a different line of attack. Instead of trying to see the "big-picture" in terms of a global theory of vision, like feature-detection, Fourier analysis or Marr's theory of vision, many researchers in biological vision have concentrated on trying to better understand the functioning at a local level in the cortex or in parts of the visual pathway. The hope is that a more precise description, possibly in mathematical terms, would constrain the functional possibilities sufficiently to construct a full scale theory. Much research work has been carried out on the parallel subdivisions of the visual pathway mentioned above; on the topographic and non-topographic maps between peripheral and cortical subsystems and within the

---

[4]For example, it has been shown in these references, that the response of "even" cells to bar stimuli can be modulated by motion of a textured background, while most complex striate cells show different responses to texture and bars. More recently it has become clear that many of the response properties and selectivities ascribed to cortical neurons are only displayed in *anaesthetized* animals. Under normal sensory conditions when the animal is awake and behaving, the stereotyped neural responses to "features" are no longer as apparent or as stable. Even such factors as posture, affect the response for unchanging visual stimulus. This and other recent evidence lends increasing support to the notion that the function of perception is not to map an environment, but to control the organisms own movement and more generally, interaction within the environment [26].

cortex; on the processes of learning and development at individual neuronal and synaptic levels; and on thoroughly describing the selectivities of individual cells to a wide variety of visual stimulus attributes. Included among the stimulus attributes investigated are various aspects of location in visual space, size, orientation, motion (direction and magnitude), colour, stereo and perspective depth, depth from shading, spatial frequency and many others.

5.3 Invariants

Much play is made of the term *"invariant"* in describing the process of perception, particularly from the semantic viewpoint. The feature detection theories relied on the detection of invariant features despite changes in illumination, pose, amount of occlusion etc. Marr [9, p.29] credited J.J. Gibson with directing debate on visual perception away from the "philosophical considerations of sense-data and the affective qualities of sensation" to the notion that the senses "are channels for perception of the real world outside". Gibson's starting point is the question: "How does one obtain constant perceptions in everyday life on the basis of continually changing sensations?" According to Marr this showed that Gibson correctly understood the problem of perception as one of recovering "valid" properties of the external world from sensory information. The properties or invariants that Gibson had in mind were "higher-order" variables like time-to-collision, the rate of change of texture density, binocular transformation, the ratio of angular height to angular width and other ratios and proportions like these.

> *These invariants correspond to permanent properties of the environment. ... The function of the brain, when looped with its perceptual organs, is not to decode signals, nor to interpret messages, nor to accept images, nor to organize the sensory input or to process data, in modern terminology. It is to seek and extract information about the environment from the flowing array of ambient energy [175].*

Also, according to Marr, he thought of the nervous system, as *"resonating"* in some way to these invariants. The adherents of this so-called "ecological" approach were content to leave further explanations of the mechanisms by which the perceptual system "resonates" to invariants to the neuroscientists. Their concern became the study of animals in their environments in an attempt to discover perceptually useful invariants to which their nervous systems might resonate.

In line with his ideas on the importance of the environment (the real external world) to perception, Gibson claimed that the starting point for visual processing is the structure of light in the *optic array* rather than the point intensities of light representing the projected image on the retina. From the optic array the higher-order variables representing the *invariant* information could, according to Gibson, be *directly* detected without ambiguity or indeterminacy. The traditional theories of perception which considered the retinal image as the primary input were plagued by ambiguity and multiple interpretations because of the impoverished nature of the 2-d retinal image. Since the time of Helmholtz, it was believed that perception required "... *inference* to supplement the supposedly impoverished nature of the flat, static retinal image. These processes of inference were held to *mediate* between retinal image and perception". [56, p.321]. According to Gibson the necessity to introduce these indirect "mediating" inferential processes only arose because of the restricted nature of the retinal image as a description of the perceptual input.

The interest in using invariants of the environment as a means to perceiving the nature of the external world is one of a number of points in common between the "traditional"[5] approach typified by Marr's computational theory and the "ecological" approach to perception introduced by Gibson. The principal point on which they differ is how the visual systems are organised to perceive the external world. In the traditional approach, objects are perceived (or "reconstructed") by piecing together primitive elements such as edges and blobs using knowledge of the external world. In the ecological approach

> *there is information to specify shape in higher-order invariants in the light, and it is not necessary, or even possible to decompose such processes into more primitive psychological operations or 'computations' ... It may be a task for physiologists to unravel the complexities of how nervous systems are attuned to such high order invariants, but the ecological psychologist need enquire no further once invariant information has been described [56].*

In any theory of perception there must be some level at which direct detection of a physical quantity takes place. For the traditional theories this is normally taken to be the function of the photoreceptors which is then regarded as an "elementary process closed to further analysis" [56, p.322]. In the sense that traditional theorists would argue that this detection of incident intensity does not involve "mediation" by

[5]The terminology is from Bruce and Green [56].

knowledge of the world, the detection process of the receptors is a perfect example of what Fodor and Pylyshyn [176] would call a *cognitively impenetrable* process. Bruce and Green point out that a biochemist would not be content with this parcellation of the function of photoreceptors. He would seek explanations in terms of isomerizations induced by the absorption of photon energy and the subsequent "tidal-wave" of biochemical and ionic transport processes which amplify the signal. In a similar way, the proponents, such as the ecologists, of the so-called "direct" theories of perception are content to regard the detection of invariants as a cognitively impenetrable process, not mediated by knowledge of the external world, and from their point of view, requiring no further explanation.

In many ways this seems to be a somewhat arbitrary delineation of where explanation ends and description begins. A neurophysiologist who succeeded in locating circuits which "resonated" to the appropriate invariants would seek a functional description of the processes that contribute to this effect. But, he might also seek to explain these processes in terms of computations on information from the external world in the spirit of Marr's computational theory. Marr's [9, p.30] and Ullman's [177] criticism of Gibson's approach is levelled at the exclusion of this algorithmic level of explanation between the ecological and physiological levels:

> *... the detection of physical invariants, like image surfaces, is exactly and precisely an information-processing problem, in modern terminology ... he [Gibson] vastly underrated the sheer difficulty of such detection. [9, p.30].*

In hindsight, it is possible to see that Marr himself was also doing his fair share of vastly underrating difficulty. On the face of it, the analysis of a problem in terms of computational theory, algorithm and representation, and finally implementation, forces a thorough effort at explanation of the "information processing" aspects of the problem at hand. It also smooths the way to engineering alternative algorithms and alternative implementations which nonetheless carry out the same overall function. However, Marr's insistence on the logical priority of the computation theory level of analysis grossly underestimates the difficulty of knowing what information, (of all the information available in a stream of visual data) is perceptually useful, and how it can be made explicit. The fact that there are subtleties involved in visual perception about which we are not yet even aware, is indicated by the limited explanatory power of current models in computer vision. Both the traditional and the ecological approaches to understanding perception, labour under the assumption that we are sufficiently sagacious to know what perception needs to do. There is a big gulf between "perception helps us to 'know' about the external world", and the activity of neural systems, and we argue below that Marr's progression of representations is untenable as an attempt to bridge that gulf. Thus knowledge of invariants discovered say by ecological optics does not help us to know how to characterize these invariants in terms of the operations that need to be implemented to extract this "structure" from the optic array. Similarly, introspective descriptions of things of which we have "constant perceptions ... on the basis of continually changing sensations" like some planar surfaces, would seem on the face of it, to be suitable candidates for invariants. But are they really primitive directly detectable invariants, and if not what are — edges, blobs, features? We are again reminded of Francis Crick's admonition about our mind deceiving us at every turn about what our brain is doing. On the other hand, assumptions of the physical invariance of things like surfaces based on substantive ideas of reality closely related to Aristotelian philosophy do not necessary imply that these things are perceptually invariant as well. I personally have great difficulty in understanding how Marr and Nishihara's [178] 3-d representation could be extended to cope with natural objects and substances such as trees and water. What are the invariants in these and other similar cases? Birds and bees can build nests in, and fly around trees. What invariants are these creatures detecting? Are they the same as each other and as the invariants of our perception? Perhaps the invariants we should be talking about are not those of the illusion of physical reality presented to our consciousness, but instead elements of the neural signalling received by the brain which remain invariant despite changes in the light pattern projected on the retina. These signals are real physical quantities which we can measure and quantify and whose statistics we can estimate. Redundancy either in an individual signal or between many simultaneous signals indicates the presence of invariants. Elimination of redundancy as in the way suggested by Field described above, and the subsequent measurement of variables in the visual data makes these invariants explicit. This approach where we can measure and quantify the variables involved seems more promising than traditional approaches to the problem.

5.4 Gabor Filtering

In the early 1980s it was realized that the dichotomy between feature and Fourier interpretations of the activity of single cortical cells was an illusory one

[179,162,180]. Under the assumption of linearity (which is usually the case for simple cells[6], and components of complex cell activity) it was realized that descriptions in terms of features and in terms of global spatial frequencies are complementary. "Selectivities in either domain imply complementary [selectivities] in the other and the crucial experimental results could be captured equally well by modest versions of either theory" [134]. This realization suggested that a fuller description of cortical cell responses should be couched in *both* spatial and spectral terms rather than in either alone. Marcelja [179] and independently Daugman [134] were the first to point out that the representation of the image in the visual cortex must involve both spatial and spatial-frequency variables in its description. This type of representation, which is intermediate between spatial sampling and Fourier analysis is exactly that described by Denis Gabor in the 1940s. Although Gabor's work (see section 4.3 above) was concerned with the communication of information, rather than information processing, or analysis or understanding. Coincidently, the inspiration for Gabor's interpretation arose from perceptual psychophysical qualities attributed to the human sensing of sound.

5.4.1 Mathematical models of cortical neural responses

As mentioned above, Gabor's analysis of space/spatial-frequency representations of communication signals is just the first one of many such representations discovered since, and just one of an infinite number of possible representations of this type [131]. Most of the interest in the vision community has centred on Gabor's approach because it lends itself well to modelling the stimulus sensitivity properties of many cortical neurons. Recently, however, Reed and Wechsler [135] used the Wigner distribution, another of Cohen's infinite class of space/spatial-frequency distributions to demonstrate gestalt organization clustering in textured images. The problem with the Wigner function and another combined position-frequency distribution, the "ambiguity function" [181] is that unlike the Gabor scheme, they are non-linear. This means that the addition of a new object in the visual field affects the representation of other objects [134].

[6]Simple cells actually show a rectified linear activity . Their response to a moving sinusoidal grating is a rectified sine wave with the neurons in their normally silent state for about half of the stimulus cycle. This property is probably related to the separate "ON" and "OFF" processing of visual data. By adding visual noise to the stimulus pattern, to increase the background activity well above the rectification level, the response of the cell is shown to be linear [179].

An alternative formulation of the signal uncertainly principle in terms of simultaneous windowing or truncation operations in space and spatial frequency was explained by Slepian *et al* in a series of classic papers [124,125,126]. Here the problem was formulated in terms of maximizing the energy in a finite spatial interval of a band-limited function (or vice versa). Like Gabor, their interpretation was in the form of a sampling theorem [167]. Wilson and Spann used feature sets based on finite versions of the functions described by Slepian and colleagues, called finite prolate spheroidal sequences (FPSS's) to describe and segment texture images. They showed that by using multiple-scale representation, the effects of uncertainty can be minimized if some assumptions are allowed about the nature of the image [168,169]. It may be that rather than any one particular mathematical model having priority over all others as *the* one "selected" by the visual system, many of these methods of analysis captures in their own way, something of the character of what processing function the visual system implements.

The Gabor scheme is to expand arbitrary signals in terms of a complete set of elementary function (the GEFs mentioned above) which simultaneously maximize location in space and spatial frequency domains. For one-dimensional signals, the GEFs occupy a minimum area in the time-frequency "information diagram" corresponding to one independent datum of information. This area can be redistributed in shape to give good localization but poor bandwidth or good spectral selectivity though poor localization but it cannot be made smaller. The important feature of the Gabor model is that it allows the freedom to vary parameters of the elementary or basis functions without losing information or adding free parameters.

Following the initial use of GEFs as models for simple cortical cells much work was done to confirm the hypothesis that experimentally determined simple-cell receptive fields could be combined together to represent a coding of visual data as a Gabor pseudo-expansion [162,182,183,184,185,71,186,134]. Apart from the fact that Gabor signals with suitably chosen parameters provide a good fit to the receptive field profiles, an important discovery was made by Pollen and Ronner in 1981. They found that pairs of nearby cells with overlapping receptive fields often had virtually identical stimulus response selectivities for parameters like orientation and spatial frequency but their phases were related in quadrature within their receptive fields. This is exactly the relationship required between harmonic components to provide a complete representation in the frequency domain of the signal amplitude in the spatial domain for a restricted region of visual space.[7] This quadrature phase relationship was demonstrated in pairs of cells even when their receptive fields did not display even or odd symmetry (That is, the receptive fields were orthogonal in the way sine and cosine functions are, with a constant relative phase, but with arbitrary absolute phase).

5.4.2 2-D Gabor functions

Daugman [134] extended the one-dimensional modelling of cortical simple cell receptive-field profiles by Gabor functions to two dimensions. This allowed further testing of the Gabor theoretical framework as a model of the overall spatial receptive field response of the linear cortical cells it also allowed an investigation of the representation of orientation sensitivity and its relationship to spatial and spectral selectivity by explicitly treating them within the same framework. Daugman shows that the extension of Gabor's 1-dimensional Schwartz inequality arguments to two dimensions allows the derivation of a 4-dimensional uncertainty principle expressing the theoretical limit of joint 2-dimensional resolution which is the product of the occupied areas in the 2-D spatial and 2-D spectral domains[8]. The 2-D GEF in the spatial domain is the product of an elliptical Gaussian profile with a complex exponential. This complex exponential consists of two real-valued components, each of which looks like a sinusoidal grating with parallel peaks and troughs. Again the pairs need to be in quadrature phase to properly calculate Gabor coefficient, though they need not be centred at the origin of the Gaussian ellipse nor aligned with its axes to be a valid GEF. It is the individual real-valued components which correspond to the filter spatial impulse response[9] or receptive field of the simple cortical cells. All GEFs

[7] A complete set of GEFs spanning information space is not orthogonal. The transform based on these basis sets of functions will not be reversible like a Fourier transform [71] without the introduction of auxiliary functions biorthogonal to the GEF [132]. Presumably this problem does not arise in biological vision systems and with appropriate spacing the GEFs can be quasi-orthogonal.

[8] The effective area of a 2-d GEF in either of the spatial or spectral domains separately is given by the product of its marginal second moments along the principal axes of the Gaussian ellipse. The modulation wave vector must be parallel to one of the principal axes of the Gaussian profile ellipse – rotating out of alignment will in general increase the effective area.

[9] The (dot) product of an input image region with a GEF gives a coefficient which is the amount of energy contained in the minimum "quantal" information volume corresponding to that GEF. It is the independent datum corresponding to the GEF.

achieve the theoretical minimum "information volume" of $1/16\pi^2$ (regardless of the values of their eight parameters) in the 4-d "information hyperspace" with x, y, u and v as its orthogonal axes.

The GEFs for 2-D signals like images, have eight degrees of freedom [134]. Two coordinates specify the location of the filter in 2-D (visual) space. Two "modulation" coordinates specify the location of the filter in 2-d frequency space where the most useful representation is the polar coordinate system of spatial frequency and preferred orientation. These four coordinates must be spanned in order to completely represent

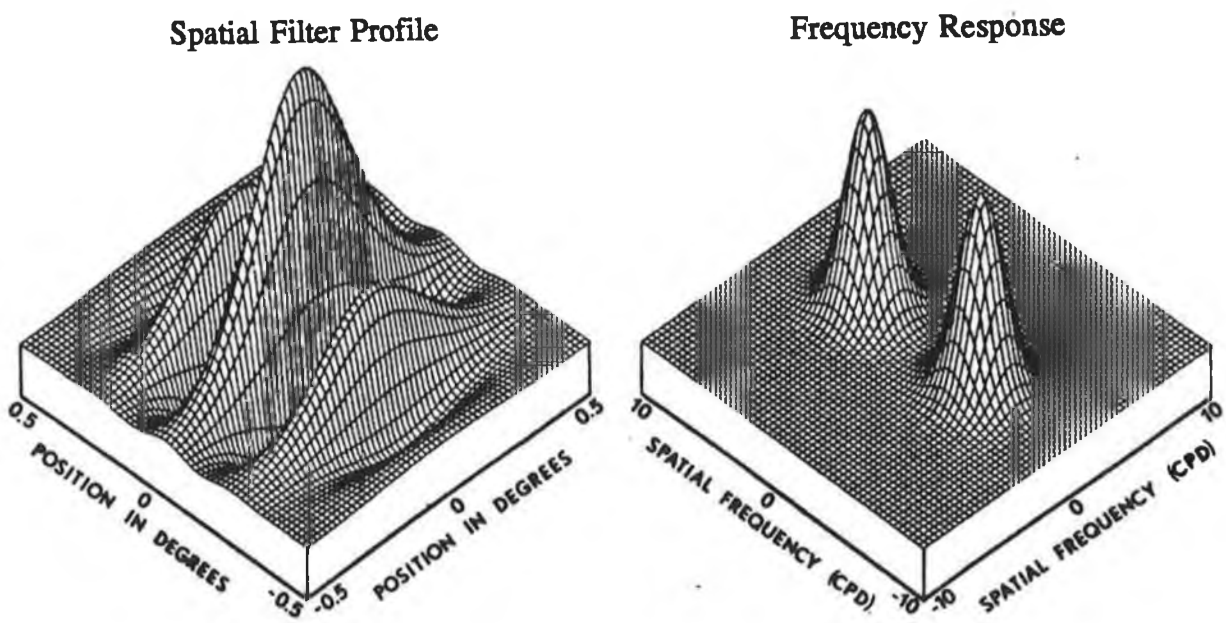


Figure 20. A perspective plot of 2-D GEF in the spatial domain (real-part) and the frequency domain. From [134].

the energy of the input image[10]. One degree of freedom specifies the phase of the modulation component. This can have an arbitrary value as long as there is another filter with identical values for all the other parameters (overlapping receptive field; identical frequency selectivity, bandwidth etc.) but with a quadrative value for this phase parameter. (There is no preferred absolute phase value, and the components do not even have to display even or odd symmetry). Finally, two parameters specify the

[10]The fact that the Gabor functions are not used to sample the image in the spatial domain as often as in the pixel representation, and do not sample the image in the frequency domain as often as in the Fourier representation, but do sample in both domains simultaneously means that there are exactly as many degrees of freedom — exactly as many Gabor coefficients — as there were pixels in the original image.

length and width of the Gaussian envelope (reciprocally related in the conjugate domains) and one specifies the angle between the principal axes of the Gaussian ellipse and the modulation wave vector orientation. They parameterize bandwidths and the "axes of separability".

In the 1-d case, Mackay [162] described how the receptive fields of simple cells (which were invariably measured as 1-d profiles at that time - 1981) represented a near-optimal solution to the problem of sampling optical images in *both* the frequency and the spatial domains. But he goes on to point out that "their relatively wide bandwidth (low "Q") suggests that at this stage the inevitable information-theoretic compromise is weighted in favour of spatial resolution". The extension to a description in the framework of 2-d GEFs shows the same trends in bandwidth (as it must) but now includes the added sophistication of allowing a division of labour which could favour either spatial resolution or orientation bandwidth, i.e., any gain in spatial-frequency resolution is offset by a loss in orientation resolution and vice versa if the filter effective area remains unchanged.

Despite the very wide variation of receptive-field dimension, orientation bandwidths and spatial-frequency bandwidths among neurons in the visual cortex, a strong positive correlation has been reported [187,188] between orientation bandwidth and spectral bandwidth in the cat striate cortex. In one case [188], the orientation half-width increased by 10° per octave increase in spatial-frequency bandwidth. This fits well with a predicted value derived from the GEF model at an aspect ratio of 0.6 [134]. Now, for 2-d GEFs occupying a fixed "information volume" in the 4-d conjugate product hyperspace[11], centred on a particular modulation frequency we must have an *inverse correlation* between orientation bandwidths. On the other hand, for 2-D GEFs with a fixed aspect ratio but occupying different areas the orientation and spectral bandwidths will be *positively correlated.* If λ is the aspect ratio (in both domains), the GEF has moderate spectral bandwidth and the modulation vector is parallel to the profile ellipse principal axis, then the positive correlation is shown by the explicit expression [134]:

[11]The 2-D equivalent of the information diagram for 1-D signals.

$$\Delta\theta_{\frac{1}{2}} = \arcsin\left[\lambda \frac{(2^{\Delta\omega} - 1)}{(2^{\Delta\omega} + 1)}\right]$$

where $\Delta\theta_{½}$ is the orientation half-bandwidth and $\Delta\omega$ is the spectral half-bandwidth (in octaves). This relation is independent of the receptive-field modulation frequency ω_0, though both orientation and spectral bandwidths can be made sharper at the expense of spatial accuracy by increasing ω_0 or increasing the size of the 2-D spatial profile.

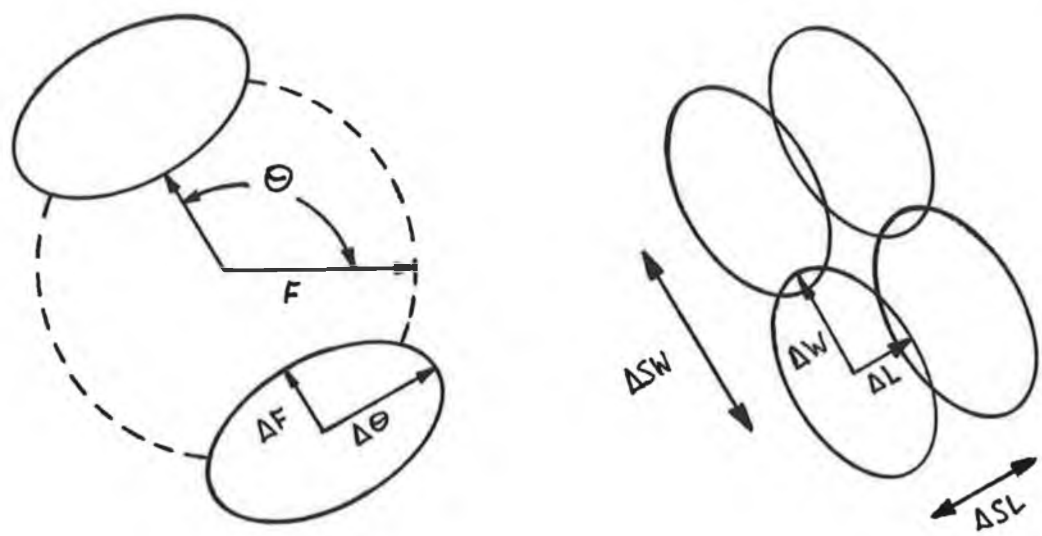


Figure 21. Schematic illustration of the relationship between the Gabor profiles in the spatial and the frequency domains. Adapted from [71].

It is interesting to examine this connection between orientation and spectral bandwidths, and other correlations between calculated parameters for GEFs fitted to cortical cells. Daugman points out that

> *... significant constraints on the degrees of freedom of the 2-D filter family ... may reveal an important underlying logic in the sampling scheme that paves information hyperspace. ... The variety of orientation bandwidths and spatial-frequency bandwidths encountered suggests that different cells occupy information space with different strategies, sometimes favouring $\Delta\theta_{½}$ at the expense of Δy, sometimes favouring Δx at the expense of $\Delta\omega$, and so on ... [representing] a division of labour among diverse strategies for extracting different kinds of spatial information subject to the inescapable uncertainly relations [134].*

Pollen and Ronner [184] discuss this latter issue in terms of whether bandwidth is reciprocally related to preferred spatial frequency or whether receptive field size is inversely proportional to the centre frequency where the bandwidth in octaves would be a constant independent of the preferred spatial frequency. They conclude from experimental evidence that neither system is solely in operation; although there is a tendency for bandwidth to narrow with increasing spatial frequency, there is considerable scatter in the bandwidth at any spatial frequency. An intermediate system with scatter and redundancy may be used; narrower tuning might be useful for periodic stimuli while broadly tuned spatially restricted fields would be more useful for scrutiny of fine detail. In the region of the cortex receiving axonal projections from the fovea, the aspect ratio λ varies between ¼ and 1. In the same region there is a 30:1 range in the diameters of simple-cell receptive-field centres – corresponding to a 1000:1 range of areas. Thus it seems that for some as yet unknown reason, the ratio b/a is constrained and relatively stable, while the product ab is either unconstrained or forced to span a large range and varies by a factor of 1000 between cells in the same region of cortex. In the same way, it seems the modulation axis is constrained to lie along the receptive field axes but there seems to be no strong constraint on the absolute modulation phase value (only on the relative phase value of a quadrature pair).

Lest one gets carried away with enthusiasm for the Gabor theoretical framework on the basis that because it is optimal in some sense, the visual system *must* implement it; it is important to remember a few caveats. Most of these results have been derived for the cat visual cortex – we have no real idea how the results will carry over to the more complex primate cortex. Not all simple cells, even in the cat, fit this model. There is some evidence that other models may better represent a larger set of cells [189]. Although the authors in this case admit to having more parameters in their model than in the 2-d GEF, they do point out ways in which the Gabor model differs from the experimental data. The main reason why GEFs do not provide a precise fit for the spatial-frequency tuning curves of cortical simple cells is because they fail to capture the relative symmetry of the cells' tuning curves on a log axis [71]. Field describes a log-Gabor function which is symmetric on a log axis and so may provide a better description. The fact that the Gabor framework, which has very strong theoretical reasons for being the processing model "of choice" in the cortex, is not exactly implemented again leads one to believe that it is not a mechanism based on some particular mathematical processing model that the visual system implements. Either it seems to indicate that the system develops in a way that may not be describable *a priori* by a theoretical model because the final operational mechanism depends on accidents of development, or whatever underlying quantities are being optimized lead towards some ideal or optimal configuration but don't succeed, or perhaps even need to get there for successful operation. At least one gets the impression that, in spite of the fact that it is difficult with the current level of knowledge to ascribe a purpose to this type of processing, of the sort envisaged by Marr as a computational theory, that the results and models are nearer to the processes that cause the system to develop than the computational theoretic descriptions.

Although thus far the Gabor ideas have only been used in a comparative, descriptive manner, the overall unified theoretical framework is a solid basis from which to begin the "*interpretive* debate" about what Gabor-like representations achieve and beyond that to constructing a new theory of perception.

5.5 A Model Gabor Code

Apart from the entropy localization properties of the Gabor representation, the approach provide no further insight into the behaviour or development of cortical neurons [71]. While the Gabor code is an efficient means of filling "information space" to provide a complete representation of any signal or image, Field points out that this does not necessarily mean that such a code is an efficient way of representing the information in any image or set of images. In fact, for a wide variety of images, a Gabor code will be quite an inefficient means of representing information. In contrast, the Karhunen-Loùve transform is *always* the optimal linear way of representing a particular ensemble of signals with minimum least-squares error. But while literally anything can happen in an image, in general, the "things" that do "happen" are drawn from a relatively small subset of all the possibilities.

> *The fact that natural images are optical projection of 3-d objects whose physical constitution and material unity yield local autocorrelation and homogeneity, whether this be merely the local luminance value or a more subtle textural signature captured by some higher order statistic, leads to anisotropic local correlations which in turn suggest appropriate sets of primitives for capturing and representing image structure. The existence of such correlations in image structure, which reduce image entropy and which can be made explicit by an appropriate set of image primitives is the*

fundamental reason that image data compression is possible. ... Images containing non-random structure or coherence, as natural images do, have a statistical complexity which does not correspond to their resolution (number of resolvable states). ... Efficient neurobiological or artificial visual systems must exploit the statistical correlations inherent in image structure. [190].

This theme of the efficiency of a processing or coding system depending on the statistics of its input was taken up David Field in a paper published in 1987 [71]. We examine Field's ideas and findings in detail because of the clear insight which they give to perceptual coding.

5.5.1 Efficient codes and perception

According to Field, an optimal or efficient method for encoding or representing information about observed scenes depends on two things:

(i) the goal of the code and

(ii) the statistics of the input.

Traditional theories of perception have tended to fall down in one or both of these areas. For example the Fourier analysis theory mentioned above seemed to have little *a priori* justification other than that it performed a type of global analysis (which must be useful), and it is a helpful mathematical tool in many other application areas. Similarly, edge detection theories were based on the not unreasonable assumption that because physical boundaries or edges are useful, their projection on an image should be perceptually useful. However, there was and still is substantial confusion between the two separate and only *vaguely* related concepts of a physical edge in the external world and an intensity "edge" in an image. The confusion is compounded by the fact that the nomenclature does not even distinguish between the two concepts. Marr's computational theory tried to make the goal of a system (in the sense of its processes and representations) explicit, but he made the error of assuming that we know enough about the world to say how it should be represented. In insisting on the top-down computational theoretic approach, Marr fell foul of Field's second criterion. It is not surprising that this should happen, because the mainstream vision community never really took perception of natural scenes very seriously.

Perhaps one reason for this tendency to concentrate on what we *think* needs to be perceived rather than what has been influencing the development of our senses throughout evolution, is the fact that much of the development of computer vision has been motivated by the application areas of robotics and automation. The objects of perception here are usually quite unlike anything in the natural world. They are often simple regular geometric shapes, easily interpreted in terms of geometric concepts such as lines, planes and volumes. (Cf., for example, the geometric cylinders representations of Marr and Nishihara [9, 191] and others). Another possible reason is the strong reductionist philosophy which permeates all fields of scientific endeavour. To our mind, what might seem conceptually like a simple or primitive stimulus might bear scant resemblance to what the visual system is optimized to see. Symptomatic of this is the concentration within the computer vision community or extracting information from "snap-shot" image stills. Nothing in the world we have grown to be able to see is ever really still – either we are moving or things within our environment are moving – and this is reflected within our visual system[12]. If an image is stabilized on the retina it *disappears* within seconds. The older magno system (which we seem to share to some extent with lower vertebrates like the cat and) which seems to be responsible for detecting movement and depth and the structure of the world around us, is particularly vulnerable in this way. Field points out that our present theories about the function of the visual cortex

> *... are based primarily on the response of such neurons to stimuli such as checkerboards, sine-wave gratings, long straight edges, and random dot patterns ... There seems to be a belief that images from the natural environment vary so widely from scene to scene that a general description would be impossible.*

He thus sets out to show that images of the natural environment show a number of consistent statistical properties which help in the interpretation of the processing of the visual system.

The goal of the system, needed to satisfy Fields' first criterion for an efficient code, comes from Horace Barlow. For thirty years Barlow has been emphasizing the need to understand visual processing in terms of the reduction in the redundancy inherent in natural images [16,192,103,104,105,193]. The purpose of natural image processing is according to Barlow

---

[12] An even stronger position on this point is that our visual systems are not "optimised" to perceive our environment at all, regardless of whether it is "natural" or geometric, but *to control our movement*. The fact that we are capable of appreciating or describing our environment may be a side-effect of the main purpose of vision, or possibly a later adaptation associated with the development of the parvo system.

to represent visual scenes by activity of a sparse selection of reliable and non-redundant (ie., independent) elements [105].

Field introduces a coding model based on the Gabor theoretical framework, in order to show how the response properties of cortical cells could be an efficient way of representing spatial visual information from the external world. There is nothing special about the Gabor framework in this context: It simply provides a tractable mathematical structure which quite accurately models the cell responses. The Gabor code has many different elements and levels which are normally distinguished in a verbose descriptive manner, and so to avoid confusion he uses the following terminology. A *sensor* is an individual Gabor elementary function with particular values for all its eight parameters. A *channel* is a spatial array of sensors tuned to the same spatial frequency and orientation. A *code* is the entire set of channels required to represent an image. The GEFs are not in general orthogonal, so there is no unique or canonical way of constructing a code or selecting the parameters for channels and sensors – as mentioned above, different combinations give different "divisions of labour". The model used by Field has characteristics which makes it compatible with, and allows it to span the parameter combinations found in mammalian vision systems.

(i) The code is chosen so that the sampling distance in space (sensor spacing) and in frequency (channel spacing) is proportional to the size of the function in space and spatial-frequency domain respectively. This means that the channel separation in the frequency domain is determined by the frequency bandwidth which determines the width of the individual sensors and the spacing along the width of the sensor.

(ii) Similarly, the orientation bandwidth specifies: the spectral distance between adjacent orientation channels; the length of the sensor (parallel to the modulation lobes, perpendicular to the modulation wave vector); and the separation between neighbouring sensors in the length direction.

(iii) The model is chosen with the sensor frequency bandwidth proportional to the sensor central frequencies. This means that the orientation bandwidth is constant in degrees and the spatial-frequency bandwidth is constant in octaves, leaving the freedom to choose the actual size of the bandwidths.

(iv) Like the pairs of cells described by Pollen and Ronner [184] with quadrature relative phases, this model contains pairs of orthogonal phase-selective sensors.

(v) The combination of sampling, bandwidths, etc., used in this model mean that the output code can represent the same number of independent data as there are input pixels. That is, the total number of sensors (each one gives an output coefficient) is constant and equals the number of free parameters in the input image which is simply the total number of input image pixels. The spatial-frequency or orientation bandwidths can be chosen without affecting these values.

This last point illustrates one of the convenient features of the Gabor framework mentioned above: it allows the freedom to vary parameters of the sensors without loosing information or adding free parameters [71].

5.5.2 Influence of statistics on perceptual codes

Despite the apparent complexity of patterns and incomprehensible range of possible images and scenes, which mostly look very different to us, Field claims, like Daugman quoted above, that they do have statistical properties in common. One of these properties is that for many natural images, the power spectrum falls off as $1/f^2$ – approximately at least. This type of power fall off is reasonable if the relative contrast energy of the image is roughly scale invariant (independent of viewing distance) as this gives equal energy in equal octave frequency ranges.[13] Because the code model being used has spatial-frequency bandwidth channels which are a constant octave width (their bandwidth increases with f measured on a linear scale), this means that, on average, the channels of the code all carry equal amount of energy. By Parseval's theorem, which relates energy in spatial and spectral domains, equal amounts of energy (in the spectral domain) carried by the various channels of the code, gives roughly equal response activity or variance for the channels in the spatial domain.[14] If the

---

[13]For a $1/f^2$ power falloff, Field quotes a fractal dimension value of 2.5 for the luminance profile of the image.

[14]Just to clarify this last point: the channels are selective for ranges of frequency in the Fourier spectrum which become relatively broader with increasing frequency. There is a corresponding selectivity in the spatial domain to patterns with ranges of feature size or detail. The range of sizes of detail selected for by a particular channel decreases (becomes more selective in range of sizes that give response) with decreasing detail size. (Proportionately equal differences between two small objects and two large objects involve much

(continued...)

assumption of stationary statistics across the entire image is made, then different sensors have roughly the same probability distributions and they thus carry approximately equivalent amounts of information *on average.*

The use of the $1/f^2$ power falloff of natural images to show that cortical cells seem to be "tuned" to carry approximately equal amounts of information (because of the constant bandwidth in octaves) is very encouraging. It shows a metabolic balance between the work rates of different cells which is compatible with the average demand made by the external world on them to carry out their diverse processing activities, measured in terms of Shannon information. As Field points out however, the constant bandwidth in octaves does not make the code more or less efficient – even the original pixels carry equal amounts of information if the image statistics are stationary. Other parameters need to be varied consonant with the constant octave bandwidth to increase the efficiency. One of these parameters is the bandwidth value itself. A constant octave bandwidth means that all the sensors have the same value of bandwidth (in octaves). But they could all be changed to some other constant value in octaves it this were to change the efficiency.

To achieve the goal set out above by Barlow (representing information by "a sparse selection of reliable and non-redundant elements") our code must satisfy a number of criteria:

(i) Its coefficients (the sensor output) should be statistically independent – or approximately so. This means that the possibility of a few sensors representing a particular image is not compromised by two sensors with strongly overlapping overall selectivity. Also, because sensors code for different (orthogonal) aspects of the input, information about these aspects is made more explicit.

(ii) Information should on average be evenly spread over the arrays of channels and sensors. This gives a better utilization of available processing resources, greatly increasing the amount and breath of processing ability over codes which make less effective use of their resources.

---

[14](...continued)
smaller absolute differences in the smaller case). Equal variances in these channels mean that differences between objects occur in proportion to the objects overall size, rather than these differences being equally distributed on an absolute scale, as is the case with white noise.

(iii) The sensors should have a large signal to noise ratio. In the case of cortical neurons which have a small dynamic range and are relatively noisy, this means that they should signal strongly or not at all. At the very least they should not be required to carry signals which depend for their interpretation on subtle changes in the activity of an already noisy component. Implicit in this is the suggestion that any *particular* image should be encoded by the *smallest* possible number of *active* sensors. The image information is thereby spread over a few sensors which are consequently very active for this particular image, (they are carrying all the "energy" for this image), with large signal-to-noise ratios.

The first two criteria are automatically satisfied by the Gabor code described. It is interesting to try to satisfy the third criterion, by varying the sensor bandwidth, which is a free parameter of this Gabor model.

5.5.3 Coding to optimize the SNR

Two extreme examples illustrate how variation in the spectral bandwidth is an effective parameter for tuning a code so that it responds in the manner described above for particular ensembles of images with different characteristics. One is the set of images which consist of sparse randomly distributed point intensities like a picture of the night sky on a starry night. The other is a set of images with obviously periodic patterns of light intensity like a picture of the weave in cloth. In the first case, which corresponds to the "spatial" extreme of the Gabor framework, little if any further coding need to be done to achieve a code which satisfies the stated criteria. The "pixel" representation already codes the visual information in a sparse array of responses (for each image), which are easily distinguished from their background (large SNR), which are on average (over the set of images) all just as likely to be active at some time or other, and finally which have little or no effect on each other (i.e., are independent). In this extreme, sensors with the widest possible spectral bandwidth (point or delta-function representation) allow most of the total variance for any particular image of this type to be represented within the smallest possible subset of sensors. A Fourier transform of an image from this sparse dot-pattern ensemble would correspond to a code with the opposite extreme of the narrowest possible bandwidth. These images would have a broad spectrum with most of the Fourier coefficients (sensor outputs) being non-zero, with values which need to be accurately represented to capture the original image information. For this ensemble of images, the Fourier code would be a grossly inefficient way of coding the image data.

The second ensemble of images corresponds to the "spatial-frequency" extreme of the Gabor framework. Here the periodic patterns, which in the spatial representation require the constant activity of virtually all the pixels, can be represented in the spectral domain by a small number of non-zero Fourier coefficients (sensor outputs). In other words, for the highly periodic ensemble of images, more of the total variance corresponding to the original image's information is represented by a smaller subset of sensors when the sensors' bandwidth is narrowest. Broadband pixel representation is the inefficient extreme in this case.

As with the two extreme cases presented by Field, the goal of a code in a system where "natural" images are the primary visual input, is to represent most of the image variance, (ie maximize the energy) within the smallest possible number of strongly active sensors, for each image.[15] By measuring the proportion of the variance represented by the most active sensors, as a function of sensor bandwidth, Field was able to show that the optimal sensor bandwidths in the sense of the efficiency of the representation of image energy (or variance, or information), are in the range 0.5 to 1.5 octaves for the natural images used. When sensors with bandwidths in this range are used to code "natural" images the maximum amount of energy is represented by the large activity of the fewest sensors. Because few sensors represent the total energy of any particular image with relatively high activity, if the sensors are noisy (as in the case of cortical neurons), they do so with the best possible signal-to-noise ratio.

5.5.4 Gabor models of cortical cells

There is some evidence that the Gabor elementary function fails to capture the precise form of the spatial-frequency tuning curves in monkey cortical cells, possibly because they lack the relative symmetry which the tuning curves often display on a log axis [71,189]. Field suggests that the log-Gabor function (the profile of which is automatically symmetric on a log axis) might provide a better description in cases of

[15]Note that different subsets of sensors will be active for different images. This means that efficient coding in this context can not involve the reduction of dimensionality to a particular subset of sensors, as is the case with the Karhunen-Loève transformation. Sooner or later each one of the sensors will be strongly active.

wider sensor bandwidth (i.e. > 1 octave). It is also more compatible with the polar distribution of sensors in the frequency domain used in his Gabor model. Once again this illustrates the point, that it is not *which* theoretical model of the processing or coding function is optimal (in our terms) that is important. It does not matter what theoretical model we use as long as it captures the "real goal" of the processing or coding system. So far the best description of the "real goal" that we have is not feature extraction or spectral estimation, but Barlow's "sparse activity of reliable and non-redundant elements". This is exactly the spirit in which Fields's analysis is carried out. The visual system during development is not trying to build Gabor or log-Gabor filters or Canny edge detectors because it "knows" from millennia of years of evolution that these are optimal is some sense. It is trying to find the best way of extracting maximum information from its environment with the available metabolic resources. This means attempting to represent information in as efficient a manner as possible, which can only be effective if the system is "tuned" to the statistical properties of this information.

Thus we see that the response properties of cortical cells, captured in this case by Gabor or log-Gabor functions in 2-dimensions, are quite well matched to the statistics of natural images. Their range of bandwidths measured in octaves show that there are probably interspersed populations of neurons optimized to carry out efficient information processing on different types of input (and therefore with different statistical properties) as there must be to support the form/motion division of labour, for example. Interestingly, the fact that the cortical coding systems seem to be matched to the statistical properties of their particular type of input, giving a representation in which only a few units are active at any time (the units "giving a large response or no response at all"), bears a resemblance to the implicit goal of feature-detection systems. The output of a feature detecting system, like the pandemonium model in cognitive psychology [107], feature-detection in mechanical pattern recognition [194], or edge-detection in computer vision [195], is also presumed to consist of the reliable activity of a small number out of a large set of independent detectors. However, Field's coding model does not describe any type of "feature-detection" in the general sense of the term normally used in the vision community. He does not attempt to determine statistics of the environment which might be biologically significant to the animal, or to give arbitrary preference to any particular object or event in the environment. Rather, "*information* is defined in relation to the *variability* of the images, not any specific feature" (emphasis added) [71]. The goal of the code (to be an efficient representation), and the statistics of the ensemble of input images, alone determine the response properties of the Gabor sensors in this model and by implication in many of the cortical simple cells. Human selected feature sets or top-down theories of what a visual system *should* be representing play no role.

5.6 Gabor Codes for Transforming Redundancy

The non-redundancy of sensor responses is one of the criteria demanded of the code model above. Measures of redundancy of various orders in the input images are also a useful means of interpreting what the cortical code is doing overall. According to Kersten [157] one of the distinguishing characteristics of intelligence is the ability to make accurate and reliable predictions or inductive inferences from partial or incomplete data. This ability "depends on the existence of statistical dependencies or redundancies in natural images". Attneave [196] like Barlow [192-193] also claimed that the principal task of biological vision is to encode the visual image in a less redundant form so that "rather than searching for features in an image, the visual system codes a given image with regard to its relation to the statistical properties of the set of natural images" [157]. The enormous number of degrees of freedom in the image signal captured by a camera or eye means that some type of dimensionality reduction is inevitable. Exploiting naturally occurring redundancy to reduce the amount of data being processed and transported makes good engineering sense. The same amount of information can then be carried by fewer neurons with a smaller dynamic range. These are exactly the arguments that we have used above to explain the processing function of the retina. But there may be even stronger reasons than efficient transport for redundancy reduction in the visual cortex. Kersten suggests that image understanding tasks may be simplified by reduction of redundancy. We argue below that perception is *only* possible because of the existence of naturally occurring redundancy in visual data. Eliminating redundancy or "detecting invariants" is a fundamental part of the process of perception.

5.6.1 Redundancy

First, we carefully define what we mean by redundancy and than discuss Field's results about efficient coding in the visual cortex in terms of the usual ideas about redundancy. The n^{th} order redundancy in an ensemble of digitized images is defined in terms of the n^{th} order conditional probabilities of the pixels. Assume for the moment that the images are ideally sampled at k points arranged in an appropriate lattice and than the sampled values quantized with an accuracy of m bits per sample. Then if the probability of grey levels is constant (all levels equally likely) and independent of all other pixel values in the image the entropy takes on a maximum value of m bits per pixel. Knowing previous or neighbouring pixel values tells us nothing about the value we could expect from a particular pixel before we look at it. This figure of m bits per pixel acts as reference against which predictability and redundancy can be measured. If there is some redundancy between this pixel and others in the image, then knowing the values of the other pixels concerned allows us to make predictions about the likelihood of the pixel having each of its possible values. In other words, we can construct a probability distribution over the possible grey levels for this pixel, $p(i)$. If there is redundancy, the probability of getting grey level i depends on the particular values of (some of) the neighbouring pixels. Suppose n neighbouring pixels affect the value of the pixel under consideration. These n pixels can appear in $(2^m)^n$ different combinations or states (labelled b_j, for $j = 1$ to 2^{mn}, say), assuming order is important. Each one of the possible 2^{mn} states that the neighbouring n pixels can take up can affect the probability of our grey-level i differently. So for each grey level i we can define a function of the set of states labelled by b_j, $p(i/b_j)$, called the conditional probability of grey-value i, given (or conditional on) the particular combination or state, b_j. In reality, this is a set of 2^{mn} different probability distributions labelled by b_j, over the grey-levels labelled by i which our particular pixel might take up when next observed. Suppose the n neighbours are found to be in a particular state, b_j, for some particular value of j. Then the probability distribution $p(i/b_j)$ allows us to work out our expected (average) ignorance of the grey-levels of this pixel on condition that the neighbours are in state b_j. This is described by the conditional entropy of the distribution over grey-levels for the pixel assuming its neighbours are in the state b_j:

$$S^n_{b_j} = -\sum_{i=1}^{2^m} p(i|b_j) \log p(i|b_j)$$

But the n neighbours are just normal pixels which can take grey values in the range 0 to $2^m - 1$, each with its own probability distribution $p_l(i)$ over the grey-levels i, where l ranges as a label from 1 to n. We can thus define a probability distribution over the

2^{mn} states, where the state b_j occurs with probability $p(b_j)$. If S_{bj} is the entropy assuming that state b_j has occurred among the neighbours then the expected value of S_{bj} over the ensemble of states of n neighbours is

$$S^n = \sum_{j=1}^{2^{mn}} p(b_j)S^n_{b_j} = -\sum_{j=1}^{2^{mn}}\sum_{i=1}^{2^m} p(i|b_j)p(b_j) \log p(i|b_j)$$

and $p(i|b_j).p(b_j) = p(i,b_j)$ which is the joint probability distribution over the $n + 1$ pixels. S^n is the n^{th} order conditional entropy for the ensemble of images. The limit as $n \rightarrow k$ is the minimum average number of bits per pixel required to code images from the ensemble.[16]

Recall that $p(i)$ is the *a priori* distribution for the pixel grey-levels before the values of any of its neighbours is known. There is a corresponding entropy function S^0 measuring our "average surprise" when the pixel value becomes known. It describes an uncertainty or ignorance about the grey-level values. $p(i|b_j)$ is the *a posteriori* distribution of probabilities when the values of n of its neighbours are tested. S^n is the corresponding *a posteriori* n^{th} order entropy measuring our remaining uncertainty about the grey-level values of our chosen pixel when the values of a particular set of n of its neighbours is known. The information that the n pixels convey to us about their common neighbour which is the subject of our interest is

$$I^n = S^0 - S^n$$

Two possible definitions of the n^{th} order redundancy could then be

$$R^n = \frac{I^n}{S^0} = 1 - \frac{S^n}{S^0}, \quad \text{or} \quad R^n = 1 - \frac{S^n}{m}$$

Kersten uses the latter, which simply assumes that all pixel grey-values are equally likely (a flat histogram) before the information about this pixel conveyed by its neighbours, is taken into account. His goal is to estimate the entropy of a set of natural images. However, the calculation of anything other than the lowest order entropies is impractical. Instead he computes bounds on the entropy and redundancy of the image set using human and algorithmic nearest-neighbour predictors to guess pixel values.

[16]Strictly, the calculations just apply to the pixel selected for discussion, but if stationary statistics are assumed the same calculations hold for the whole image.

Both the human and median predictors gave upper bounds on the entropy for the set of images used, of around 1.3 to 1.4 bits per pixel.

First order statistics concern the probability distribution $p(i)$ of the grey-levels of a pixel without any knowledge of its neighbours. The grey-values are mutually exclusive "events" which can be indexed by a finite subset of the natural numbers. Each probability assignment $p(i)$ is a number drawn from the reals which satisfies

$$p(i) \geq 0, \quad \forall i; \qquad \sum_{i=1}^{2^m} p(i) \;=\; 1$$

The natural representation of this probability distribution is a 1-d array of real numbers indexed by i, or graphically as a block histogram indexed by i in its natural order.

5.6.2 Transforming from high to low order redundancy

There is redundancy in the first order statistics when the pixel grey-levels are not uniformly distributed – when some intensities are more likely to occur then others. this introduces some degree of predictability to the intensity values. The amount of predictability is quantified by the first order entropy function S^0 above. Second order statistics concern the probability distribution of a pixel's value conditional on knowledge of the value of one of its neighbours. They are completely specified by the joint probability measure defined on the product space of the two individual sets of experimental outcomes. (This looks like a 2-d histogram labelled by the grey-level indices of the two pixels in this context). Unfortunately, second order redundancy cannot be seen by simply looking for non-uniformity in the 2-d probability distribution. It depends on the extent to which the 2-d distribution cannot be separated into the product of two 1-d distributions. A spatial auto-correlation function computed on the basis of pixel values being values of a "discrete-time" random process contains the complete second-order redundancy information about any pair of pixels.

Kersten [157] has shown that even third order entropy, S^2, does not capture all the predictability that can be achieved by humans or a simple nearest neighbour median predictor. There typically is redundancy in the values of three and more neighbouring pixels that is exploited by humans, but not captured by the power spectrum and autocorrelation functions. One of the stated objectives of Fields's code model is that the sensor responses be non-redundant. Depending on the overlap of the GEFs in space and spatial frequency, this condition is approximately fulfilled by the quasi-orthogonal functions. But the code model which Field describes contains as many sensors (64K) as there are pixels in the original image, so the overall redundancy of the code must be exactly the same as the overall redundancy of the image ensemble. A reduction in redundancy would allow the set of images to be coded with fewer total independent data – reducing the dimensionality – which clearly does not happen. The problem is not with a superfluity of sensors channels. The full complement of sensors is required to completely cover the information space defined by the size and bandwidth of the image. To omit any sensors or increase the relative spacing would result in the loss of information as well as data. It would seem, that despite the semi-orthogonality between individual sensors, we are no nearer Barlow's goal of reducing overall redundancy.

Fortunately the code, which is previously described as efficient because of the high signal-to-noise ratio and paucity of simultaneously active elements, can also be described as efficient because of its substantial *transformation* of the data redundancy. In the original image ensemble, the overall redundancy is spread over many orders – the more one knows about the nearest 4, 8 or 16 neighbours and sometimes about pixels much further away, the more predictable is the value of the pixel under consideration. But the GEFs (sensors) of Field's Gabor-code model are quasi-independent. Having knowledge about the values of the responses of neighbouring sensors, or indeed any other sensors in the model, hardly makes the response value of the sensor of interest any more predictable This means that between the output values of the Gabor code there is very little 2^{nd}, 3^{rd} or higher order redundancy. Since the overall redundancy has remained unchanged, the Gabor code model must have the effect of converting high-order redundancy in the input images into large first order redundancy of the output code. First-order statistics describe the probability distribution of the grey-levels of a pixel or the probability distribution of the response-*values* of a sensor. Large first-order redundancy means that the sensor responses are very unevenly distributed with some values occurring very often (for example, very large values or very small values), and some very seldom (intermediate values). This was the case with the spatial representation of the sparse point intensity image and with the spectral representation of the largely periodic image, and now with the Gabor code representation of natural images if the code bandwidth is chosen suitably.

> *With a one octave bandwidth, the information [in a set of natural images] is packed into the smallest number of sensors, giving a highly skewed distribution and therefore a redundant code. In other words, the most efficient code by our terminology is the code with the most redundant first-order statistics [71].*

Overall redundancy has not been reduced by this type of representation. It remains for further stages of processing to make the most efficient use of the outputs with these first-order statistics by coding only the non-redundant elements (in other words, the highly active sensors), or whatever other processing is relevant to the continued existence of the organism. Since any of the sensors can be highly active at some time or another, subsequent processes cannot simply ignore the output of some sensors, as happens by definition with the KLT. Perhaps Barlow's goal should be altered to the necessity of reducing the redundancy of the code at any instant.

The point about ignoring the output of any sensors is important here. The Karhunen-Loûve transform, which is based on coding in terms of the largest eigenvectors of the covariance matrix of the input data, is an efficient code in the sense that it minimizes the mean-square-error (MSE) for any given data reduction. The eigenvectors of the covariance matrix are automatically fully orthogonal. They can also be arranged in order of the amount of energy they represent in the input image ensemble by the size of the eigenvalues (which are the square-root of the variances of the ensemble along the corresponding eigenvector). The KLT works by making the information carried by the complement of sensors as *unevenly* distributed as possible and eliminating those which account for small amounts of the information (or energy, or variance).[17] (Recall how the Gabor code model distributes the information as *evenly* as possible between the complement of sensors).

[17]A point of note here is that perceptually important information is often contained in low-amplitude high frequency components of image data. Because the KLT removes the lowest-energy components, the transform often causes a level of perceived distortion which is belied by the contribution to mean-square-error. This effect could be overcome by transforming the data in a perceptually relevant way, so that equal amounts of energy in different coefficients are equally perceptually relevant.

5.7 Summary

Like Marr, our aim is to try to find the answers to the "What?" and "Why?" questions about visual perception. Unlike Marr we believe that the key to answering these questions lies, not in a computational theory for vision, but in the three crucial notions: information, development and measurement. The notion of development is an undercurrent to the logic of our approach which is dealt with explicitly in the previous chapter in the section on Linsker's Infomax, but plays a much more background role in this one. Measurement is something which is taken up in the next chapter and the following one. In this chapter the development of the role of information theoretic concepts in understanding visual perception problems, begun in the last chapter, is continued.

The notion of cortical cells as feature detectors is first discussed and discarded as untenable. At the level of each cell there may a localized filtering function but the belief here is that a more general description in information theoretic terms is more appropriate. Certainly there is no basis for any sort of Fourier-analysis type model. Even though the information theoretic notions go a long way to helping us to understand the role of parts of the cortex, it is difficult to be as precise as in the retina because effects are much more dispersed and much less is known about the system. Certainly the form/motion or parvo/magno division and the macro-structure of the cortex seem to be related in some way to the structures that we see in Gabor codes but we cannot yet be sure of the details, or why. The notions of a statistical redundancy interpretation of Gabor codes allows useful insights into the possible functions of these codes and gives hints to possible applications not related to vision.

Chapter 6

6 Pattern Recognition, Concepts and Generalization

6.1 Introduction

Pattern recognition as it is commonly understood is only one aspect of human visual capacities. One of the primary points that we are trying to make in this thesis is that at some level of description, all aspects of perception can be at least approximately modelled by something like primitive pattern recognition events. The philosophical justification in chapter 2 and the information theoretic descriptions in the subsequent chapters have all been more or less leading in this direction. In this chapter we examine more closely what pattern recognition is, including different ways of representing the data to be classified and the logico-algebraic implications of these. This in turn leads to a discussion of Watanabe's propensity theory and the basic properties of observations or measurements.

Watanabe's 1985 book [19] is devoted to discussion of many different aspects of pattern recognition in humans and machines. The title of his first chapter is "Pattern Recognition as Seeing-One-In-Many" – a phrase which very neatly captures the two very different but complementary aspects of pattern recognition. On the one hand, we have pattern recognition meaning "seeing as" a member of a class, i.e., the process of being able to collectively refer to many different objects[1] by one name (or symbol) – a process of *generalization*. On the other hand we have pattern recognition meaning "seeing as" a shape, where one object is seen to be formed from a collection of many parts – a process of *grouping* or gestalt organization. Watanabe traces the relationship between the two aspects:

> *a pattern is the opposite of chaos; it is an entity, vaguely defined that could be given a name ... when we see a [pattern] we are somehow associating the present particular case before our eyes to other cases of [patterns with the same name], ... recognising similarity [19, p.2].*

---

[1]The Aristotelian view of the world as consisting of a discrete number of self-identical objects with a relatively fixed set of properties or predicates is a tacit assumption of modern mechanical pattern recognition. We have already argued against this view but continue to use it in this discussion because it is in these terms that the ideas are normally formulated.

The ability to recognize similarity allows us to recognize or identify an object as a member of a family or class. It allows us to group objects together. This clustering habit – characteristic of intelligence – may have grown out of adaptation to primitive generalizations: similar causes have similar effects. Whatever its origin, man in particular seems to have a natural instinct for recognizing groups and making a classification of his own. The word "cognition" is sometimes used to describe the formation of new classes – the process of "clustering", ie., "taking cognizance of the existence of a group of similar objects". In this terminology, "re-cognition" is reserved for the process of identifying an object as a member of an already known class. In general usage of the term "recognition", this distinction is not made and the word covers both cases.

6.2 Patterns and Classes

The two complementary aspects of pattern recognition mentioned above involve two complementary uses of the term pattern. In the former a pattern is an object *qua* a sample of a class. In the latter a pattern is an individual object as a form, shape or figure, whole or gestalt[2]. In other words the term pattern "includes not only the one imitated but also those which imitate". In mechanical pattern recognition, the emphasis is on patterns as samples of a class with two variants, one involving this notion of imitation. The *epicentric concept* involves the notion of a "central object,[3] which is the master of other imitations" [p.7]. An *aggregative concept* involves relationships (e.g. similarity) between class members but without a "typical case, exemplar, or archetype" – without a central object or imitated.

A concept, or its corresponding class is defined in logic by its *extension* or its *intension*. The extension (or denotation) is the set of all particular objects making up a class and therefore corresponding to the associated concept. The intension (or connotation) is the set of predicates which is just sufficient to identify all the objects in a class and differentiate them from objects outside the class, i.e., just sufficient to define a concept. Normally however, in real life, neither the extension nor intension are available. Even if they were it is likely that it would be completely impractical to

---

[2]This use of the term pattern is associated below with mathematical notions of entropy and structure.

[3]Compare Aristotle's adaptation of Plato's Form.

use either. Instead, intelligence seems to be able to generate and use concepts with enormous, and often ill-defined, intensions and extensions, by means of a few class samples or paradigms[4] (and sometimes samples from outside the class or negative paradigms). "Theoretically [or logically] speaking this is not a definition of a class, but it generates a working capability of distinguishing a member from a non-member of a class".

Mechanical pattern recognition does not deal with concepts or classes in terms of the logical concepts of intension and extension. Rather, like human pattern recognition, the aim is to be able to show a few paradigms (and their class affiliation) in a learning phase, and subsequently be able to classify an object (whose properties, but not class, is known), in a decision phase. This is the essential function of both human and mechanical pattern recognition: inferring a general concept from a few concrete cases. It is an *inductive* inference, with no necessary or logical basis. In mechanical pattern recognition, extra-logical, extra-evidential heuristic measures of similarity are often introduced to overcome the inductive ambiguity and allow the classification to proceed by logical computations. These heuristics either implicitly or explicitly express the value judgement of the computer programmer through the weight we attach to each variable, or threshold which we set, etc. While the heuristic principles have a role to play in resolving the inductive ambiguity of generalization, they are not foolproof. They can only be judged by the usefulness of the resulting classification. As well as this "paradigmatic" pattern recognition, there is pattern recognition in the sense of clustering or forming new classes (or cognition). This involves further inductive ambiguity which is discussed below.

> *We cannot condemn some of the enthusiastic experts of pattern recognition claiming that all intellectual acts are pattern recognition. Indeed, inductive inference and pattern recognition are basically identical in their essential function – inferring a generality from a few concrete cases [19, p.9].*

Whatever about all intellectual acts, the central thesis of this document is that perception involves, at a certain level of description, both of the complementary aspects of pattern recognition ("seeing-one-in-many"). Perception we submit, can be described in terms of many acts of primitive observation or classification or

---

[4]The word "paradigm" which Watanabe uses in this context, is in the sense of its meaning a pattern, example or model – a sample of a class. See footnote 1, chapter 2.

measurement, and also in terms of grouping processes which are capable of associating the activity of the primitive observations at different levels of abstraction, often in parallel pathways. Watanabe's comparison implies that our theory is at least along the right lines, so we examine this relationship between perception and the formal theory of pattern recognition more closely. But first we discuss the role that sensory experience plays in the development of a perceptual system.

6.3 The Basis for Perception in Experience

In the discussion on the controversy of the Universals above, we concluded that certain ideas borrowed from Platonic philosophy capture what we believe is a better understanding of the nature of perception than heretofore. The most important is the emphasis on the primacy of properties or predicates over substance: particular objects can only be identified by the observation of predicates – i.e., "testing the applicability of general concepts". (We define these terms carefully below). Since we know nothing about the world, except what is perceived via the senses, everything we know must be based on relationships between different observations. We see what we do see, not simply because that is the way the world is, or because we are born with the ability to see in this way but because of a lifetime of experience of constructing perceptions from sensations.

To tease out a little better what we mean by this we consider the example of learning a natural language. Prelingual infants have an innate ability to learn language. In fact, we can make an even stronger statement than this. Children have an innate ability to *construct* language. When a heterogeneous group of adults with no common language are thrown together by force of circumstances, they develop a pidgin in order to communicate. This is not a true language with grammatical structure and syntax. The dictionary describes it as "a jargon incorporating the vocabulary of different languages" – a hotchpotch of words separated from their function as nouns, verbs, adjectives etc., in their own language and used in a virtually arbitrary fashion. It is a makeshift language with a limited vocabulary and with almost non-existent structure and syntax, and few prepositions [197, p.179ff]. What is astounding though is what happens in the first generation after the original mixing of cultures and languages:

> *The adults who create pidgin speech are not able to provide it with any structure; they're past the critical age at which syntax develops. The*

children, however, are not. Syntax develops in them just as naturally as any other part of their bodies. It's natural, it's instinctive, and you cannot stop them doing it. I think the only explanation you can have for the way syntax works is that, somehow, it is built into the hard wiring of the brain [198].

The children in these circumstances create a creole language. This is not a language with a history, and literary or oral tradition, fashioned over thousands of years. It is a proper language with well-defined vocabulary and a relatively complete grammatical structure. It is a completely new language constructed by children in one generation. As further evidence that people have in their brains the inherent machinery to make language, grammatical similarities have been found among creoles spoken all across the world.

Notice that we have said: brains have the machinery *to make* language – not that brains have the inherent machinery of language. Chomsky claimed that

linguists should search for 'language universals' – the similarities among all languages ... The capacity for language is uniquely human and is not learned through experience ... but is innate [197, p.178].

However, the search for true universal similarities in all languages has been singularly unsuccessful and the notion of a "universal grammar that determines much of the surface structure" is without support. There are further indications of what is happening in the brain during the development of language. At birth, any child anywhere in the world is capable of learning to speak any language. By eight months, a child will have lost the capacity to make or to distinguish some sounds of other languages which are not present in its mother-tongue. By the end of a sensitive period lasting until the age of seven or eight years a child must have been exposed to some kind of language if a true language is ever to be learnt. Children seem to innately have the mechanisms that allow a language capacity to develop, but the language capacity itself is *not* innate as Chomsky claimed. The mechanisms for developing syntax generation and comprehension modules within the brain require language *experience* in order to carry out their task, and this experience must occur during a critical phase lasting up to seven or eight years. It is experience that determines the exact path that the development will take, and the final linguistic ability. Yet this experience, as evidenced by creole development, need not even be a proper language. Any type of communicative interaction is sufficient to trigger and guide the development of true linguistic capacity. Children do not simply have neuronal mechanisms which allow them to *learn* to *imitate* language. They can actively construct their own concepts about their environment, their cultural milieu and about communication and its structure, based on (often) indirect experience from, or about, any of these.

It is notoriously difficult and often misleading to argue about other functions or capabilities in the brain purely on the basis of linguistic concepts. Two particular reasons for this first came to light in work inspired by Gazzaniga's findings on "split-brain" patients. The two hemispheres seem to have very different modus operandii in their interaction with the external world. Loosely speaking, the left hemisphere is verbal, sequential, temporal, logical, analytic and rational, while the right hemisphere is non-verbal, visuo-spatial gestalt, synthesising and intuitive.[5] [19, p.465; 197, p.188, 26]. Care is needed in describing visual function in terms of language related activities because (i) at an abstract level there is a very fundamental difference in the way information processing is carried out and (ii) because our linguistic description from one hemisphere of what is happening in visual processing in the other hemisphere bears little resemblance to what really seems to be happening. According to Crick [57] "we are deceived at every level by our introspection". Nevertheless, the general characteristics of development in vision and language in children do show similar characteristics. In both, experience is a very important factor in determining the eventual cognitive ability. The existence of a true mother-tongue will determine whether a child will learn its mother language or develop a completely new creole. Similarly, if the visual data available to a developing bird is arranged to be devoid of horizontal information, it has been shown to cause the bird to come to "see" vertical information only and to ignore all horizontal information – never to regain this aspect of vision which is common to all properly developed birds. Equally, both show a critical or sensitive period, during which the child or animal must be open to experience in order for the capacity to develop normally. Finally it seems that both have initial genetically-specified information processing structures which allow (or indeed drive the young towards) the development of neuronal structures which directly subserve adult capacities. In particular, we see as we do because our minds have the right structure for us to use appropriate experience to develop adult ability. Just as

[5]Primary sensory and motor capabilities are located in each hemisphere for the opposite side of the body. So, for example, each hemisphere processes visual information in an identical fashion to the other – at least in the early stages that are known about. It is really only high level cognitive abilities, with a locus relatively remote from the primary areas, that show the traits listed here.

there is no unique universal pre-language or syntax which is a foundation for the learning of all languages, so there is no necessary view of the world which is instrumental in causing visual abilities to operate in a certain way or perception to involve one particular way of looking at things. Just as language, even sign-language cannot develop in the absence of attempts to communicate, so vision cannot develop in the absence of appropriate visual experience.

This last comparison hides a difference between ways of attributing ability to experience in the two cognitive capacities singled out. While statistical methods have been used in linguistics for various purposes, the author is not aware of a statistical *explanation* or interpretation of the development of language processing, nor how this might be attempted. On the other hand as Laughlin [15] and Field [71] have shown, statistical or information/communication theoretic interpretations of the form and function of information processing in the retina and cortex are very successful in explaining the experimental observations. These explanatory power of information-based ideas may even be further extended by the semantic considerations of particular messages described by Dretske. Barlow [192,193] has been saying for years that interpretations of visual processing in organism and computer in terms of redundancy is much more appropriate than whatever conceptual theory like edge/feature detection or geometric representations was then in vogue. The nature of sensory modalities with their well-defined physical input means that ideas like "sampling" and "entropy" are easily applied. With language processing, the cortical regions of interest, Broca's area and Wernicke's area are relatively far (in synaptic terms) from the direct, measurable, input and output contact with the external world. Unlike the peripheral visual input to the cortex it is very difficult to quantify the variables involved and relate them to statistically describable stimuli.

We have suggested, following Platonic ideas described by Watanabe [19] that what we perceive is based solely on the relationships between different observations: the nature of our perception is determined by a life-time of visual experience. This is a bit like Locke and Berkeley's ideas about perception being constructed through a process of learning through association. It is also a bit like the "transactional functionalism" idea that "what one sees will be what one expects to see, given one's life-time of perceptual experience" [56, p.92]. One significant difference between these philosophical or psychological positions and the model presented here based on psychophysical and neuroscientific evidence is the following. The implicit assumption in the former positions is that the type of process involved in the relationship between observations is describable at an abstract or cognitive level similar to the language based tokens and procedures normally associated with A.I. programming in Lisp or Prolog. In the latter an attempt is made to explicitly define statistical relationships between observations in terms of redundancy and Shannon-information content. Another point of note is that virtually all the development leading to the mature perceptual capacity takes place within a number of relatively short critical periods early in a child's life. The onset of any of these critical periods like that associated with the development of binocular vision coincides with the first arrival of neuron axons growing from more peripheral parts of the perceptual system. The critical period usually lasts for several weeks or months and thereafter little further long-term adaptation – certainly on this scale – occurs again. This burst of development followed by a life-time of relatively quiescent application is difficult to reconcile with an ongoing process of perceptual construction by association. In visuo-semantic terms the range of perceptual experience available to an infant in the first weeks and months of its life is usually very restricted. In fact there is some suggestion that early parts of the visual pathway develop *in utero* in ways that do depend on the neuronal signals being carried even when these signals arise from random fluctuations in receptors without any external perceptual input. The burst of development *is* compatible with interpretations of the mechanisms driving development in terms of statistics and information theory. Despite the limited semantic range of early perceptual experience it is expected that the range of statistical variation is as extensive as the total ensemble of possible images of all possible scenes. In fact, the existence of a critical sensitive period during which most development takes place, followed by a stable period of application may be a necessary part of some as yet unknown aspect of the mechanism of perception. It may even be a necessary aspect of the development of an ontogenetically flexible but perceptually stable system.

6.4 Inductive Ambiguity in Pattern Recognition

6.4.1 Supervised pattern recognition

Consider the probability measure $p(H_j \cap A_k \cap D_i)$ defined on the product space $\{H_j\} \times \{A_k\} \times \{D_i\}$, where $\{H_j\}$ is the set of hypotheses available, $\{A_k\}$ is the set of auxiliary conditions and $\{D_i\}$ is the set of possible outcomes. According to Watanabe there are three types of inductive inference:

(i) Given D_i and A_k we are interested in guessing the right H_j, i.e. in evaluating the inverse conditional probability $p(H_j \mid D_i \cap A_k)$.

(ii) given D_i and H_j we want to derive $p(A_k \mid D_i \cap H_j)$.

(iii) given only D_i we want to derive $p(A_k \cap H_j \mid D_i)$.

Pattern recognition involves two stages of inductive inference of different kinds, and extra-evidential, non-necessary factors are involved in each stage.

Suppose we have two classes labelled by $k = 1,2$ where A_1 states that "object O belongs to class 1" and A_2 states that "object O belongs to class 2". Suppose that H_j states that each member of class 1 has probability $f^j_1(D_i)$ of giving experimental result (a representation vector) D_i and each member of class 2 has probability $f^j_2(D_i)$ of giving experimental result D_i.

The deductive probability of object O having experimental value D_i is then

$$
\begin{aligned}
p(D_i) &= \sum_j \sum_k p(D_i|A_k \cap H_j)p(A_k \cap h_j) \\
&= \sum_j (f^j_1(D_i)p(A_1) + f^j_2(D_i)p(A_2))p(H_j)
\end{aligned}
$$

If A is given,

$$p_j(D_i) = \sum_j p_k(D_i|H_j)p(H_j)$$

During the *training period* of a pattern recognition process, the classifier is shown a D_i with its class affiliation A_k. The training process is one of evaluating the hypothesis H_j using the Bayesian rule

$$p_k(H_j|D_i) = \frac{p_k(D_i|H_j)p_k(H_j)}{\sum_l p_k(D_i|H_l)p_k(H_l)} = \frac{p_k(D_i|H_j)p_k(H_j)}{p_k(D_i)}$$

and improving the quality of the credibility through the experimental facts $D_i \cap A_k$. This improvement process is a sequential one involving the substitution of the *a* *posteriori* probability of one stage for the *a priori* probability of the next stage. As the number of stages increases, the influence of the *a priori* probability of H_j on the a posteriori probability of H_j decreases, though insofar as the evidence is finite, the prior probability can always overcome the evidential factor.

During the *application period* of the process, the classifier is given D_i and asked for the probability of its class affiliation A_k. That is, it is required to evaluate (for the second type of inductive inference mentioned above)

$$p(A_k|D_i) = \frac{\sum_j p(D_i|A_k \cap H_j)p(H_j)p(A_k)}{\sum_j \sum_k p(D_i|A_k \cap H_j)p(H_j)p(A_k)}$$

Usually, e.g. in a single neural network, only one H_j is taken as true at any one time so we use

$$p(A_k|D_i) = \frac{p_j(D_i|A_k)p(A_k)}{\sum_k p_j(D_i|A_k)p(A_k)}$$

An example where many hypotheses would be current and under evaluation at any given stage would be the application of a genetic algorithm to a population of artificial neural nets with different internal structures and processes.

Usually the prior probability of class $p(A_k)$ is taken from the actual relative frequency of members of the various classes but this might not be justifiable. For example, the known and unknown class samples may be taken from different populations. So, quite apart from the problems associated with reliably evaluating probability density functions in situations where there are only a small sample of class elements available, the process of pattern recognition intrinsically involves at least two different sources of inductive ambiguity of different kinds.

6.4.2 Clustering

Usually in pattern recognition, the number of classes and some paradigms (class-samples) are given during the training period, and objects without class-assignment are given during the application stage. The task of the classifier is to place these latter objects into the classes exemplified by the paradigms. The number of classes and a set $\{H_j\}$ of possible hypothesis, i.e. possible statistical definitions of classes, are available to the classifier. In the clustering problem, neither of these things are known. Clustering is the process of grouping objects into classes where only the properties of each object are available – no information is available on possible class assignment of these objects. It is assumed that the members assigned to any particular cluster are in some sense "bonded" together more intensely than members of different subsets. The task of deciding how many clusters there should be and statistically defining the classes or clusters corresponds to Pierce's "abduction" and hence involves further inductive ambiguity above and beyond that involved in supervised pattern recognition. As well as these theoretical difficulties there are practical difficulties associated with the entirely arbitrary measurement variables available, and, if the classification is based on similarity, the measure of similarity chosen (i.e. the metric in the data space).

During the training period of a pattern recognition task, the inductive probability of the first kind $p(H_j \mid D_i \cap A_k)$ is determined. If each object is represented by an n-component vector x, then D_i is a point in the corresponding n-dimensional space. A_k designates the class to which this point belongs. In many problems we do not need to know the exact value of $p(H_j \mid D_i \cap A_k)$, but only want to obtain the particular H_j that would maximise this. This is the approach of parametric methods of pattern recognition.

In the application stage we are concerned with the second kind of inductive probability

$$\sum_j p(A_k|H_j \cap D_i)p(H_j)$$

or simply

$$p_j(A_k|D_i) \quad = \quad p(A_k|H_j \cap D_i)$$

using only the H_j with the current highest credibility. Here D_i is the vector of the new object whose class affiliation is yet to be decided upon.

Suppose there are two class, A_1 and A_2. We would expect to classify an object x in class 1 if

$$f(x) = \frac{p(A_1|H \cap x)}{p(A_2|H \cap x)} > 1$$

However, if "misclassifying" an object from class 1 into class 2 causes a greater loss than vice versa, then we could classify an object to class 1 only if $f(x) > \theta$ for $\theta < 1$. Now θ can be determined if the loss function is known. The border surface or "decision function" $f(x) - \theta = 0$ depends on the loss which in turn depends on the usage of the classification.

6.5 The Geometric Representation of Perceptual Data

We have already mentioned above the two different interpretations that are usually attributed to "information". When it is said that there is "information" about a scene contained in an image of that scene, the term is being used in a sense very different from the classical Shannon view of information which is related to probability and counting alternatives. This is like the differentiation described by Bossomaier and Snyder between form and statistical redundancy. They contend that the processing of form information, which is a multi-level task, is "greatly assisted by removing statistical redundancy and producing an economical representation at each level" [155]. In the case of a standard computer-stored video image with 512^2 8-bit pixels, individual pictures can be represented as points in a 2^{18} dimensional space with each dimension quantized to 256 possible values or positions. Of this enormous set, only a small fraction would be recognized by people as being interpretable in any way. The remainder would in whole or in part look like random noise. Is it that these images have some *intrinsic* property that allows an attempt by humans at interpretation, or is it simply a tautological position that we can recognise or interpret the individual pictures that "make sense", i.e. those that we can recognize or interpret?[6]. Compare for example, the evolutionary tautology: the fittest are those that survive.

[6]Remember that from a computer's "point of view", all images stored in its memory, no matter how visually realistic to us, "look" as unintelligible as random noise does to us.

It seems clear that small changes in the intensity of individual pixels would not greatly affect the interpretability of most images, nor would changes in the overall average intensity. In the geometric representation, these translate into images corresponding to nearby points in the data hyperspace, looking very similar, or images corresponding to vectors with identical directions, looking very similar. The fact that interpretable images are redundant means that they can be represented by fewer bits (or equivalently that they correspond to a proper subset of the signal space), and still carry the same information for us.

Consider an ensemble of N objects on each of which we measure n observable properties or predicates. This gives a total of nN quantities:

$$X^{\alpha}_{i}, \qquad \alpha = 1,\dots, N; \quad i = 1,\dots, n,$$

which is referred to as the *object-predicate table*. Following Watanabe [19, Appendix 3] we can consider these either as N vectors with n components, or as n vectors with N components. The first case is a geometric representation of the Aristotelian view of the world as consisting of a countable number of particular objects with a fixed set of well-defined attributes. The second case is a geometric representation where the predicates are considered as the primary quantities and rather than each number being considered as the extent to which each predicate in turn holds for a particular object, the numbers can be interpreted as the extent to which each object in turn satisfies (or has a measured value of) each predicate. This latter viewpoint has been referred to as *"object-predicate inversion"* and was introduced by Watanabe as a Platonic reaction to the Aristotelian world view, which is so ingrained in mechanical pattern recognition and computer vision.

6.5.1 Object-predicate inversion and the covariance matrix

Normally the covariance matrix for an ensemble of objects is calculated as the expected value over the ensemble of the outer product of the feature vectors corresponding to objects in the ensemble:

$$C_{ij} = \sum_{\alpha=1}^{N} X^{\alpha}_{i} X^{\alpha}_{j} \qquad \dim[C] = n \times n.$$

The object-predicate inversion viewpoint gives a natural definition of an alternative "covariance matrix" that can be calculated. This is the outer product of "object vectors" for each feature:

$$D^{\alpha\beta} = \sum_{i=1}^{n} X^{\alpha}_{i} X^{\beta}_{i} \qquad \dim[D] = N \times N.$$

In fact, Watanabe shows that the eigenvalues of both of these covariance matrices are identical, and the author has pointed out the relationship between these ideas related to the KLT and the singular value decomposition (SVD). The identical eigenvalues of the two covariance matrices correspond to the (square of the) singular values of the original data considered as an $N \times n$ matrix, under the Singular Value Decomposition[7]. This relationship is discussed in more detail in chapter 8.

Continuing the theme of the geometric interpretation of the object-predicate data, the outer-product of any vector with itself has very particular properties. If $\underline{A}$, $\underline{B}$, and $\underline{C}$ are any three vectors then the *outer* or *dyad* product of $\underline{A}$ and $\underline{B}$ (often just written as juxtaposition) is defined by the following equation:

$$(\underline{A}\,\underline{B}).\underline{C} = \underline{A}(\underline{B}.\underline{C})$$

Thus the outer product of a vector with itself is a linear operator with the following properties:

(i) it produces a vector whose length is the dot-product of itself and the vector it operates on (to the right).

(ii) the direction of this new output vector is the direction of the vector whose outer product was taken.

That is, the outer product of a vector with itself produces a *projection operator* onto this 1-D vector subspace defined by the vector. This means that the covariance matrix

[7]Sirovitch and Kirby [222] use an algorithm equivalent to this system of object-predicate inversion to apply the Karhunen-Loève transform to the coding of images of faces in registration. In fact, while the diagonalization of the dual covariance matrix and the singular value decomposition are equivalent mathematical procedures, from a purely numerical computation point of view, the SVD introduces lower order errors.

is the sum of the all the projection operators corresponding to each vector in the ensemble.

6.5.2 Geometric representations and their algebraic structure

There are two important classes of data vectors which were briefly mentioned above, that are conveniently described in the geometric representation. One class consists of all vectors where the *actual* value of each component in the vector is important and is captured by the notion of the "volume picture". The other class consists of all vectors where the *relative* values of components are important. This second class corresponds to data from such things as sampled speech or images. Here the absolute magnitude of the vector component is measured or transduced, but is not relevant for subsequent interpretation. The natural representation for the first class of vectors is an n-dimensional *metric space* with the Euclidean metric. The natural representation for the second class of vectors is a *vector space* where all the vectors are of unit length. If we consider each component of a vector as the result of evaluating the property corresponding to that component on the object represented by the vector, then the vectors corresponding to a collection of objects can be arranged into an *object-predicate* table, the components of which we label by the matrix of elements $\{X^{\alpha}_{i}\}$ above.

Now, the principal conclusion drawn from the discussion on realism in chapter 2 was that the mind can only apprehend the particular by virtue of it being able to apprehend universals. It was based on a belief in the need to subvert the notion of substance which held a primary position in Aristotle's philosophy[8]. Consider the following, more explicit statement of this idea in answer to the question of how an object (particular) is identified:

> *It is identified by observation, just as a predicate is confirmed or denied by observation. A raven is identified by first observing all kinds of predicates that ravens are supposed to satisfy, and then by observing some special marks of this-ness, such as the one that lives in a certain tree, or the one that has a defective right wing or some other characteristic. ... We mentioned a recent work by Strawson, which maintains that a particular*

[8]Watanabe compares this subversion of substance implicit in the object-predicate inversion with the negation of substance (anatman) in Buddhism. A similar shift from substance to function is noted in the introduction of the quantum theory of elementary particles where the self-identity of elementary particles must be relinquished.

> *object can only be identified through testing applicability of some general concepts (universals), which, in our context [pattern recognition], amounts to observation of some predicates. If we agree that an object can be identifiable only by a group of observations, the object-predicate relation is no more than a relation between two groups of observations. [19, p.92]*

Now, consider the following extract from C.S. Peirce which further supports this view and extends its implications even into logic:

> *I have maintained since 1867 that there is one primary and fundamental logical relation, that is illation ... A proposition, for me, is but an argument divested of the assertiveness of its premise and conclusion. That makes every proposition a conditional proposition at bottom ... This is the very same relation that we express when we say that 'every man is mortal,' or 'men are exclusively mortal.' For this is to say, 'Take anything whatever, M, then if M is a man it follows necessarily that M is mortal.' [199]*

These ideas supply us with an indication of how to represent predicates in the geometric representation of pattern recognition data [19, p.510]. A proposition $P(a)$ that an object a satisfies predicate P means, according to the usual Aristotelian interpretation of pattern recognition, that object a is placed in class P; (the class of all elements for which P is true). The interpretation of $P(a)$ suggested by Peirce, is that if x satisfies A, which is the predicate or property of being a (i.e. A-ness), then x satisfies P. In other words the Aristotelian logical formula $P(a)$ becomes an implication between predicates: $A \rightarrow P$. Peirce goes further though, in the first part of the extract above, claiming that the relation which underlies all logic is *implication.* So for our geometric representation of predicates, we need to find some way of representing the predicates such that the implication relation is also represented geometrically. In turn then, representations of all other logical connectives in geometric form can be derived from the basic geometric representation of implication. This means that extensive logical formulas involving our pattern recognition primitives (observables) can be decided by referring to the solution in the corresponding geometric picture, and in fact even the underlying structure of the logic can be derived from the geometric picture. What we find is that different types of pattern recognition situation (e.g. comparing absolute or relative data vector component values) give rise to different geometric representations of predicates and implication and even to different logical or algebraic structure underlying the interactions of these terms.

In the situation where the value of each component has meaning, which is the so-called "volume" picture, a predicate A is represented by a subset of the metric space (usually a volume):

$$V_A = \{ x / A(x)\}$$

The corresponding geometric representation of implication, is the volume (or set) inclusion relation: *"is a subset of"*:

$$A \rightarrow B \;\Rightarrow\; V_A \subseteq V_B$$ [9]

Watanabe shows that this representation satisfies all the laws of Boolean Algebra, including the distributive law, which is not surprising as the distributive law is also characteristic of set theory. There is a one-to-one correspondence between Boolean algebra and set theory.

In the case where only the relative values of the vector components corresponding to a single object are meaningful, each object is represented by the direction of a unit vector in the vector space; i.e. by a 1-dimensional subspace of the space. The subspace corresponding to a predicate A is then:

$$M_A = \{ x \;/ A(x)\}$$

where if two vectors satisfy A, any linear combination of the two vectors satisfy A. The corresponding geometric representation of implication is the subspace inclusion relation: *"is a subspace of"*:

$$A \rightarrow B \;\Rightarrow\; M_A \subseteq M_B$$

The predicate $\varnothing$ that represents the constant absurdity has no member and corresponds to the zero vector:

$$M_\varnothing = \{x \;/ \varnothing(x)\} = 0.$$

The predicate $\square$ that represents the constant truth is satisfied by every object and therefore is represented by the entire vector space:

$$M_\square = \{x \;/ \square(x)\} = \{x\} \;\Rightarrow\; \varnothing \subseteq A \subseteq \square$$

[9] It is important to distinguish relations belonging to the "object language" and relations belonging to the "meta-language". Consider a collection of predicates $A, B, C, \ldots$, each of which can be true or false [47, p.2]. An operation which combines members of the collection to produce a member of the collection is a part of the object language. This corresponds to using the predicate A, regardless of its truth or falsehood in a particular case. But, when we assert that A, in a particular case, is true, this assertion is not itself a predicate – we are talking *about* the predicate A and so this assertion belongs to the meta-language. Here $\rightarrow$ and $=$ belong to the meta language. $\cap$, $\cup$ and $\neg$ belong to the object language. Note the difference between A in $A \rightarrow B$ and A in $A \cap B$. The former is in the meta-language and means "the proposition that the predicate A is true". The latter is in the object language and simply means "the predicate A is true".

Using the definition of implication $\rightarrow$ we can define conjunction $A \cap B$ and disjunction $A \cup B$ for any two predicates A and B. (They correspond to the *meet* and *join* operations in lattice theory respectively).

$$A \cap B \rightarrow A;$$

$$A \cap B \rightarrow B;$$

If $X \rightarrow A$ and $X \rightarrow B$, then $X \rightarrow A \cap B$.

Thus the subspace corresponding to the conjunction $A \cap B$ is the largest subspace which is a subspace of A *and* is a subspace of B.

$$A \rightarrow A \cup B;$$

$$B \rightarrow A \cup B;$$

If $A \rightarrow X$ and $B \rightarrow X$, then $A \cup B \rightarrow X$

$M_{A \cap B}$ is the set of all vectors that can be expressed as linear combinations of vectors belonging to M_A or M_B or both. (Note that the combination can produce a new vector which does not belong to either M_A or M_B). The following laws still hold:

Idempotent Law:	$A \cap A = A, A \cup A = A.$
Commutative Law:	$A \cap B = B \cap A, \; A \cup B = B \cup A.$
Associative Law:	$(A \cap B) \cap C = A \cap (B \cap C)$
	$(A \cup B) \cup C = A \cup (B \cup C).$
Absorptive Law:	$(A \cap B) \cup A = A$
	$(A \cup B) \cap A = A$
	$\varnothing \cap A = \varnothing, \; \varnothing \cup A = A$
	$\Box \cap A = A, \; \Box \cup A = \Box$

However, the distributive law breaks down when C is not a subspace of A or B:

$$(A \cap B) \cup C \neq (A \cup C) \cap (B \cup C)$$

$$(A \cup B) \cap C \neq (A \cap C) \cup (B \cap C)$$

For finite dimensional vector spaces a less restrictive modular law holds:

$$(A \cap B) \cup C = A \cap (B \cup C)$$

which is equivalent to the distributive law when $C \rightarrow A$ because $A \cup C \rightarrow A$.

Another case where the distributive law holds in the subspace picture is if the predicates A and B are compatible.[10] This is the case, for example, if there is a rectangular coordinate system in the vector space, and the subspace in the geometric representation of any predicate, is a subspace subtended by some of the coordinate axes [19, p.514]. In general, the distributive law holds in the subspace picture when predicates share a common coordinate system.

The subspace picture, which is derived on the basis of the relative values of the vector components of the vector representing a particular object, being the meaningful quantities, shows that in certain cases the distributive law does not hold for incompatible predicates. This is equivalent to saying that the order in which predicates are measured is important and changing the order can give a different result.

6.5.3 Propensity theory

Watanabe gives another more fundamental derivation of this non-distributive or quantum logic which shows a direct relationship to the problem of measurement. Recall again the essence of Pierce's idea about the fundamental basis of logic being the notion of implication, i.e., instead of saying $P(a)$ is the proposition that the object a satisfies the predicate P, we say that if x satisfies A, which is the predicate of being a (i.e. a-ness or a-hood; cf. doghood) then it satisfies P, i.e.:

$$P(a) \text{ becomes } A \rightarrow P$$

However, continuing to make use of the notion of subjective probability described earlier, Watanabe points out that in human pattern recognition there seldom is a definite (yes/no) implication. We can seldom say that an implication is definitely true or definitely false. Usually there is some sort of "a graduated evaluation of the veracity of an implication", i.e. there is a probability associated with whether or not the implication holds. This is of the form of a conditionality probability: $p(P|A)$ which assigns a measure to the probability of P being true or applicable given that A is known to be true or applicable.

[10]The terminology comes from quantum mechanics and means that it is possible to simultaneously measure both predicates to an arbitrary accuracy, unconstrained by uncertainty relations. This is *not* the case with position and momentum in quantum mechanics or equivalently with position and spatial frequency in signal theory (see discussion on Gabor above)

There are two ways of incorporating this conditional probability, which encodes the veracity of a logical implication, into a logic system. The conventional approach is to assume Boolean or Aristotelian logic as the starting point and to incorporate notions of probability into this system. This is the approach developed by Kolmogorov, where probability is defined as a measure on a Borel field which satisfies certain properties.[11] Watanabe shows how this definition of probability is only consistent if the underlying lattice satisfies the distributive law [19, p.518]. He credits Louis de Broglie with pointing out that the use of probability is anomalous in quantum mechanics. This is because [47] the distributive law does not hold for observables in quantum mechanics[12] while the use of the usual Kolmogorov concept of probability *requires* the distributive law to hold.

One way of getting around this problem is to relax the definition of probability. The usual definition of probability requires:

(i) $p(A) \geq 0$;

(ii) $p(\varnothing) = 0$;

(iii) $p(\Box) = 1$;

(iv) If $A \cap B = \varnothing$ then $p(A \cup B) = p(A) + p(B)$.

In the distributive case (iv) is equivalent to (iv)′:

(iv)' $p(A) + p(B) = p(A \cap B) + p(A \cup B)$

To cope with a non-distributive lattice, replace (iv) by (iv)″:

(iv)″ $p(A) + p(\neg A) = 1$

which is a restricted version of (iv)' when B is $\neg A$.

In order to interpret the conditional probability $P(A/B)$ consider the case $P(A/x)$ where x stands for B when the geometric representation of B is a 1-dimensional subspace.

---

[11]Borel fields form a lattice which differs from a Boolean lattice only in the fact that a countably infinite conjunction and disjunction is allowed in the case of Borel fields. This does not materially change the results. [19, p.517].

[12]In quantum mechanics observables can be represented by operators in a Hilbert space or as matrices in Dirac's bra/ket form. The fact that operators corresponding to incompatible observables do not commute (the order of operation is important) is a mathematical representation of the fundamental nature of quantum systems. This non-commutativity leads directly to the breakdown of the distributive law.

Define $$p(A/x) = |Ax|^2/|x|$$

where Ax is the projection of x on the subspace corresponding to A.

$$p(A/x) + p(\neg A/x) = 1$$

by the properties of subspaces [19, p.519]. If $P(A/B)$ is considered as the average over all vectors x contained in B then $p(A/B) + p(\neg A/B) = 1$, which is a weaker form of the conditional probability version of (iv) above.

This problem with the distributive law needing to hold in the case of a probability defined on a Borel field can be avoided in another way. While the conventional approach assumes Boolean logic and then defines probability on top of this, an alternative approach starts with the conditional probability and uses this to derive a logical structure.[13] Watanabe claims that although the orthodox way of introducing probability is to add it to an existing logical structure,

> *it seems to be the opposite of the natural order of development of ideas in human cognition ... In ordinary thinking, a vaguely conceived association between cause and effect with a graduated degree of certainty is generated first in mind, and in rare occasions it is crystallized as an infallible implicational law. The logical axioms can be considered as a formalization of such exceptional cases of singular associations [19, p.520].*

Recall again that our starting point is the implication $A(x) \rightarrow P(x)$, and we want to introduce some way of dealing with "a graduated evaluation of the veracity" of the implication. According to Watanabe the most natural way of doing this is to allow a continuous range for the truth value of $A(x)$ or $P(x)$ which usually have one of the dichotomous values 0 or 1. To represent this he introduces a function $f(A,x)$ (for A say), such that

$$0 \leq f(A,x) \leq 1$$

where as usual the value 1 means that the object x definitely satisfies the predicate A (is in class A), and the value 0 means that the object x definitely does not satisfy the predicate A (is outside class A). The class A can be understood as the *extension* of the predicate A (the set of all objects that satisfy A). When f is limited to the values zero and one we get the usual Boolean logic out of the formalism. This is equivalent to the assumption that at any instant each predicate corresponds one-to-one to a well-defined,

---

[13]Recall the discussion above about the conditional probability being a more fundamental concept than the unconditional (absolute) probability. This discussion is on the basis of the conditional probability being a more fundamental concept than logic which is derived from it.

fixed set of objects that satisfy the predicate (i.e. a fixed extension). This is an assumption which Watanabe calls "the postulate of definite (or fixed) truth set" and which he attributes to Frege with the name "the Frege Principle", [19, p.521; 46, p.408].

This introduction of *f(A,x)* as a graduated measure of the certainty of an object *x* being a member of a class defined by the predicate *A* arose out of the fundamental role assigned to *implication* by Pierce, and the extension of this to ideas about the vagueness of human thinking by Watanabe. Having defined the *f*-function, there are two interpretations that can be assigned to it, depending on the interpretation of implication. In the first case *f* is called a *"propensity"* function. Here we assume that we have an empirical method of determining whether or not the object *x* satisfies the predicate, with the *f*-function expressing the "degree of expectation of obtaining the positive empirical result in the *A*-ness test". But, the critical point is that after the *observation* is made, the result is *either* definitely **true** or definitely **false**. After the observation any uncertainty about the membership of *x* in the *A*-set or class is unambiguously removed. In the second interpretation of the formalism, *f* is called a *"fuzzy"* function or "membership" function. In this interpretation there is no empirical method or test that can affect the values of the *f*-function. The fuzzy function expresses a purely subjective evaluation of the *A*-ness property of object *x*. We are not concerned further with this fuzzy set theory here [200]. The propensity theory, however, is extremely interesting from the point of view of perception. Using an interpretation of the process of perception introduced by Wilson *et al* [201,202] we are able to explain several interesting aspects of perception using the propensity theory.

6.5.4 The properties of measurement

Three assumptions underlie the propensity theory [19, p.521ff]:

(i) An observational method called an *A*-ness test or *A*-test can be defined to determine whether or not an object *x* satisfies predicate *A*.

(ii) The observer has a degree of expectation, *f*(*A*,x) of getting an affirmative result in the *A*-test of *x*.

(iii) The result of two consecutive tests, an *A*-test and a *B*-test may depend on the order of the two observations.

The fact that the result of two tests depends on the order in which they were carried out could be because:

(a) the observed object is changed due to the observation, or

(b) the observer changes as a result of an observation on an object, or

(c) both observer and observed change as a result of an observation [19, p.522].

The quantal nature of microscopic physical systems seems to arise from reason (a). The effect of the measuring apparatus on the physical system being measured (and therefore interacting with the measuring apparatus) has been discovered to be finite. This means that the effect of measurement on the measured system cannot be ignored, which was one of the basic assumptions of classical mechanics. Watanabe [47, sections 5.3, 5.4] discussing the fact that *information loss* is an inevitable consequence of observation or measurement, indicates that the term measurement is somewhat of a misnomer – the actual process is something more akin to preparation. Our knowledge about the system before the act of "observation" is entirely *probabilistic* and random. Our knowledge about the system after the act of "observation" is that the system is in a *definite* state which can be represented by the eigenvector (of the measurement operator) whose corresponding eigenvalue was the outcome of the act of "observation". An example of the need for a propensity theory on the basis of reason (b) is possibly the psychology of medical diagnosis: "when A and B are very close or similar to each other, the ordinary human doctor will tend to classify a patient with a higher probability into A when A is considered before B than when B is considered before A". [19, p.522].

We are generalizing the predicate A used above from the simple dichotomous case of a predicate defining a set or its set-complement, to the polychotomous case of a predicate classifying objects into one of several mutually exclusive sets - a partition. This does not affect the underlying theory, as every finite multi-way classification can in theory be reduced to a finite number of binary dichotomies.

6.6 Summary

The basic generalization process involved in generating non-overlapping equivalence classes and labelling them, is common to classification, measurement theory and pattern recognition. In chapter 2 we related this to perception itself, inspired by the notion of a universal due to Plato. In this chapter the properties of classification or pattern recognition are described in more detail, particularly the sources of the various inductive ambiguities that are involved in the process. The properties of various types of classification are discussed using the idea of a geometric representation of sampled data and we find that the properties of different types of geometry are reflected in the corresponding type of pattern recognition problem, In particular, for problems where the relative values of predicates are important (e.g. contrast), rather than absolute values, we find that the representation in terms of subspaces means that the distributive law is not guaranteed to hold. This means that it is sometimes not possible to simultaneously evaluate some (conjugate) predicates. The actual process of evaluating a predicate (measurement) also comes under scrutiny and the reasons why different predicates might not be compatible are mentioned. Because probability is based on Borel sets, it is not strictly correct to use probability in cases where the distributive law does not hold. The propensity theory introduced by Watanabe is described as a means of overcoming this problem. This chapter finally brings us to the stage where we can begin to meaningfully discuss what it is that we mean by an observation or measurement, the implications of this in terms of the breakdown of Boolean logic, and the relationship to perception in general.

Chapter 7

7 Perception as Measurement

7.1 Introduction

In a book published in 1978 on the "Fundamentals of Measurement and Representation of Natural Systems" [20], Rosen attempts to clarify the relationships between

(i) the theory of *measurement*, which forms the basis for all our knowledge of physical properties;

(ii) the theory of *recognition* mechanisms in biology and engineering;

(iii) the theory of *discrimination* which deals with the specificity of interactions between systems, and

(iv) the theory of *classification* as used in the establishment of diverse taxonomies.

A common factor in all these topics is the generation of some sort of *invariants* such as numbers, which serve to label the processes with which they are associated. This happens in such a way that processes which are considered to be alike bear the same label, and those which are considered different bear different labels. This definition of the basic element which connects these related processes prompts a number of questions which Rosen attempts to address [p.x]:

(1) What has been learned about a system when we have measured it, or recognised it, or classified it, or otherwise labelled it in a particular way?

(2) How is this knowledge related to that obtained from a different procedure for measuring, recognising or classifying it?

(3) What does it mean to say that *distinct* systems (which can be distinguished according to some criterion), nonetheless bear the same label (and are thus indistinguishable according to some other criterion)?

(4) When does indistinguishability with respect to one type of observation entail indistinguishability with respect to some other type of observation?

The approach he uses to begin to confront these questions is based on the idea that every recognition, measurement, discrimination or classification process depends on the capacity of a given system S to induce a dynamics (in other words, a change of state) in another system M, which he refers to as a meter, discriminator, recogniser or classifier.

I have been constantly and continually confronted with these very questions, in a variety of guises, in the course of my investigations into the organisation, development and evolution of biological systems, for as long as I can remember. Over the course of time I have come to realize that these questions are not merely subsidiary matters to be dealt with cursorily within the confines of a specific investigation as circumstances indicate; rather they are the ***primary*** *questions on which the resolution of all the other questions essentially depend [p.x].*

In chapter 2 above we discussed the problem of the status of universals and how one aspect of the position associated with Plato is that the basis of our knowledge about the external world is not the direct perception of real objects but the ability to apprehend universals. In chapter 6 the notion of what it means to "apprehend universals" is made more explicit: it is an *evaluation* of whether or not a particular concept, or value of a property, or predicate, applies in a given case; it is something we could represent as a number, as a component in a particular component position of a vector. According to Peirce, the most fundamental relation of logic, and therefore of all relations, is the implication relation between predicates. Furthermore, by representing this implication relation with different interpretative models (set-based, or vector-everbased) corresponding to different possible applications, we find that it is possible to have very different logical or algebraic formulations for the association between the predicates of our systems. One of these corresponds to the usual Boolean framework. Another is described which does not, and there are likely to be yet other formulations. The aspect of the problem concerning whether or not a particular predicate or value is applicable – abstractly depicted as a *test* – was discussed, and some of its properties described. We thus suggest that Rosen is correct in emphasising the measurement problem as the primary problem of the nature of the interaction of a system with its world, and therefore the primary problem of the nature of perception.

7.1.1 The Explanation of Perception

At various points we have argued that what perception does, is not *defined by* the external world, in the sense that we think of a world full of objects, events and other perceiving beings like ourselves. Neither is perception *independent* of an external world, for such a process would either be useless as an aspect of interaction, or accidental and therefore without explanation. Perception is compatible with physics and with, at the very least, the statistics of the medium of our existence, i.e., the statistics of what we call signals or physical properties. As observers, we can describe the structure of other living organisms, as having various capacities to interact with what we describe as their environment. The parts we describe as having the specific and primary function of mediating the ingoing signals we refer to as perceptual (sub-)systems. Nevertheless, *our* descriptions and explanations of the functional role of these parts of the organism cannot be a part of the organisation and structure of that system, because our functional descriptions necessarily invoke what we see as the organism's environment. That environment is something that only exists to us. In particular, our conception of that environment may have little or no relevance to the organism concerned. Our explanations of that organism's perception in terms of representations of its environment might be useful for us, but cannot be relevant to the organism, because they (the explanations) again involve our conception of that environment.

So, what *is* relevant to the organism? Well, the physical properties of the medium in which it exists, the signals mediating its interaction with that medium, its organisation and structure which determine how these signals affect it, and what, if any relevance they have. There are two points here: the seemingly arbitrary character of what is observable or what is relevant, and the fact that what *is* relevant is not determined by a single objective world, but by the current state of the organism and the history of interactions that, compatible with its *world* and maintaining it viability, have led to this state. The first point, the choice of observables and the relationship between them are the issues that Rosen is concerned with in the four questions he poses above. But the second point, the appropriate level of explanation of a living organism is a more fundamental issue and therefore something that needs to be discussed first.

Now the type of explanation at a particular level is tightly bound up with the level of explanation, so that for example, a description in terms of the dynamics of a complex system will also entail a particular "outlook" on the interaction of the system with it's environment. Thus for example, in a discussion on a systems dynamics, one would not expect to reference to 3-D representations. If then, the dynamical system can do without 3-D representations, they may be less necessary than would seem on initial consideration. Consider for example, how the following description of an "information-processing" system in dynamical terms actually reverses many of the accepted ideas on the relationship between perceptual systems, the environment, and information:

> *Perception does not begin with causal impact on receptors; it begins within the organism with internally generated (self-organised) neural activity that, by re-afference, lays the ground for processing of future receptor input. In the absence of such activity, receptor stimulation does not lead to any observable changes in neural dynamics in the brain. It is the brain itself that creates the conditions for perception by generating activity patterns that determine what receptor activity will count for it. Perception is interaction initiated by the organism, not a reaction caused by the object at the receptor level. Thus the story of perception cannot be told simply in terms of feed-forward causation in which the object initiates neural changes leading to an internal perceptual state. What is missing in the reflex-based model is recognition of the role played by self-organised neural processes and by dense feedback among subsystems in the brain that allow the organism to initiate interaction with its environment [39].*

The picture that is emerging from the work of Freeman [37, 40], Zak [41], Grossberg [203], Rosen [12] and others, is that what we describe as the functional properties of a perceptual system, are a direct result of the dynamical interactions of the components of that system, which collectively compose its structure. According to Maturana and Varela, this exactly is the appropriate level of explanation of the systems operation, because it is defined completely in terms of the systems organisation, and its realisation as a structure of components and relations between components. The system's world perturbs this structure and the system tries to maintain its identity, but these perturbations are not part of the definition of the system's organisation [24].

The work to understand the details of the dynamics of these systems is in its infancy. It seems though, that the dynamics may contain stable attractor states where the system goes into a period of relatively stable oscillation. Under the influence of external perturbation, the dynamics can very quickly switch from one stable global activity consisting of a distributed group of neurons oscillating in synchronisation, to another attractor and a completely different distributed pattern of activity. The precise nature of the attractors, particular the distributed pattern of activity, seems to be a function of many factors including the previous experience of the system, its state of arousal, its posture and so on. Loosely, however the system could be described as having a "fast" dynamics (the actual moment-to-moment collective and distributed neural activity), and a "slow" dynamics, involving the gradual change of attractor configuration and characteristics. This latter "slow" dynamics has, more correctly, been referred to as a *metadynamics* by Varela and others [53], because it involves, not just the change in the operating parameters or operating conditions of a dynamical system, but the actual change of the system itself. A description in terms of a particular Newtonian dynamical model would be slowly invalidated as the metadynamics transforms it into a different system[1].

Nevertheless, while this level of description in terms of dynamics is the only level that can be operational for the system, the system can be described at other levels. The central thesis of this dissertation is that information theoretic (IT) and measurement theoretic (MT) terms, are an appropriate level at which to usefully describe a perceptual system, even if ultimately any proper artificial realization will involve designing a dynamical system. This is the case, because this level of description displays many of the properties that are implicit in the organisation of the system, despite being at a more accessible, better understood level of description. On the other hand, because a description in terms of information/measurement is nearer to the level of operation of the system, it is less influenced by our particular prejudices on what observables are appropriate, than a purely functional description in the spirit of Marr's computational theory would be.

So, having identified an information and measurement type of analysis as an important level of explanation, we can go further to see what particular properties of a perceptual system this viewpoint lays bare. We do not attempt to describe a full theory of perception. That would require other levels of explanation in addition to this – particularly the dynamical type just considered – and perhaps other properties that are not included in the informational or measurement ideas. Instead we try to identify some of the elements or components, out of which an IT/MT theory would be constructed. The key idea is the notion of a transition from "signals to symbols" as a measurement, or classification, or primitive observation.

---

[1]This is one of the points that Rosen [12] makes in arguing against the standard reductionist approach to describing *complex* systems, which is what he claims biological organisms are.

7.1.2 Redundancy and inferences

Having carefully defined what we mean by measurement in chapter 6 and clarified the nature of our level of explanation above, we can apply this notion to attempt to explain what happens in a perceptual system. We examine the notion of a transition from signals to symbols and the role of symbols. The active or dynamic aspect of perception also is discussed.

Several authors have emphasized the fact that we can only see because of structure or correlations in natural images.

> *Our own ability to interpret the images that our eyes receive involves making inferences about the environmental causes of image intensities, often from incomplete data. This ability to make predictions or inferences depends on the existence of statistical dependencies or redundancies in natural images [157].*

Attneave's [204] and Barlow's description of biological vision as encoding the visual image in a less redundant form carried the implication that eliminating redundancy automatically makes the structure contained in the image clearer, by separately representing some of the more important aspects of this structure. The Principle of Prägnanz which embraces such properties as regularity, symmetry and simplicity was introduced by the gestalt school of psychologists to capture the notion of form-discerning. It has much in common with the heuristic principle of least entropy or the principle of simplicity [19]. All of these heuristic principles point to the existence of correlations between different parts of images when the image is of the real world. It is these correlations at many different statistical levels, which when appropriately distributed, allow an image to be interpreted by a biological vision system. The immediate questions is: what are these correlations or redundancies of the various orders, and why are images which contain these, automatically interpretable? That is, how does a visual system know exactly what correlations are important and how to use them.

One type of statistical information available in natural images is captured by the spectral power function. This indicates, as described by Field [71] that there is a $1/f^2$ power falloff with increasing spatial frequency in natural images. This is a very simple example, but it still confirms (in this case through Field's multi-equal-energy-channel hypothesis) that even low order correlation can have a profound effect on the structure of the visual system. So, according to this argument, we can *see* because our visual systems have adapted to the structure (redundancy, interdependence) contained in the ensemble of visual data of natural scenes and described by various orders of statistics. If our visual system was exposed to an ensemble with a completely different set of statistical properties, one would expect it to adapt to be able to interpret the images of this ensemble. Images of our natural world would then "look-like" random noise or little better. Unfortunately, while the idea is quite plausible, the author is not aware of a direct explanation which links the notion of redundancy and that of inference, other than in a very imprecise way.

7.2 A Semantic Theory of Information

7.2.1 Information theory and semantics

So far we have separately discussed the ideas of information theory and measurement theory. There is a direct relationship between them: the methods of information theory can be used to characterize the amount of generalization taking place at the classification or *measurement* of each input signal. The equivocation measures the amount of information lost in going from distinguishable inputs, to a single undifferentiated output signal, usually represented as an abstract symbol. Dretske's theory, which we discuss here, covers this aspect of the relationship between information and measurement, but also tries to interpret a theory of information in a semantic context.

The usual understanding of the mathematical theory of information, (or communication theory) is that it tells us something about information — if not *what* it is then at least *how much* of it there is [22, p.40]. On the other hand there have been arguments that information theory actually tells us *nothing* about information as such, rather that it is a theory of, or about *signal transmission*, i.e., a theory about the physical events that, in some sense not defined within information theory, *carry* information. According to this argument, in other words, a distinction is made between signals — abstract descriptions of physical events or properties, particularly time-varying ones — and the information they carry, their *semantic* content. Thus *information theory* is a sophisticated theory for describing the statistical features of the physical events (signals) that are used by us for communication, and their interdependence or correlations[2]; but *information* has to do with *what* exactly we communicate using these physical events or signals. A proper theory of information should thus be a theory about the content of our messages, as well as the form of the messages in which this content is embodied – in other words it should be a *semantic* theory as well as a quantitative one.

Now, according to Shannon as illustrated in the extract in section 4.1.1, the semantic aspects of communication are irrelevant to the engineering problem of describing and designing communication systems. Nevertheless, as Dretske points out, this does not mean that the engineering (quantitative) aspects are necessarily irrelevant to the semantic aspects; (see for example, Weaver's introduction to [121]). While information theory does not pretend to say anything about what specific information a signal carries, it does tell us *how much* information a particular signal carries and therefore places a *constraint* on what information a signal *can* carry. This is the basis of Dretske's attempt to use information theory as the framework on which to built a proper *semantic* theory of information, which information theory does not claim to be.

The first problem is to decide exactly what is meant by the term "semantic" in this context. Weaver, for example, cautions that in a semantic theory of information, the term information in this semantic sense should not be confused with *meaning*, and similarly it should not be confused with the *value of the information* or with *knowledge* [22, p.41]. While information in the term's ordinary usage does have a semantic connotation, this is not the same as meaning, "for, on the face of it there is no reason to think that every meaningful sign must carry information or, if it does, that the information it carries must be identical to its meaning" [p.42]. Usually people communicate (i.e. exchange information) by employing the customary meaning of signals or signs, but the two ideas are not synonymous: signals may *have* a meaning by convention and agreement, but they *carry* information. Exactly what this information that signals or signs carry, is not completely clear. In ordinary usage the

[2]Dretske makes it clear [22, chap.1] that there does not necessarily even need to be a *causal* relationship between events for the communication of information between them. In fact the two extreme cases of having full information without causality and having no information with causality are both possible. "From a theoretical point of view, however, the communication channel may be thought of as simply the set of dependency relations between [a source] *s* and [a receiver] *r*. If the *statistical* relations defining equivocation and noise between *s* and *r* are appropriate, then there is a channel between these two points and information passes between them, even if there is no direct physical link joining *s* with *r*." [p.38].

term information has different interpretations and connotations, but Dretske claims that there is a common nucleus to these different interpretations and that this can be captured in the notion of *truth.*

> *What information a signal carries is what it is capable of 'telling' us, telling us* ***truly****, about another state of affairs. Roughly speaking, information is that commodity capable of yielding knowledge, and what information a signal carries is what we can learn from it. If everything I say to you is false, then I have given you no information. At least I am giving you no information of the kind I proported to be giving. If you happen to know (on other grounds) that what I am saying is false, then you may nonetheless get information, information about* ***me*** *(I am lying), from what I say, but you will not get the information corresponding to the conventional meaning of what I say. [p.44].*
>
> *Information is what is capable of yielding knowledge, and since knowledge requires truth, information requires it also. [p.45]*

So what information a signal carries is what we can learn from it, and there need be no correspondence between the *meaning* of the symbols and the *information* conveyed. From this point of view, "*false* information and *mis*-information are not kinds of information — any more than decoy ducks and rubber ducks are kinds of ducks"[3].

The key to using the quantitative viewpoint of information theory to tell us something about the semantic aspect of information, is that while information is a commodity that is capable of yielding knowledge, given the right recipient, exactly what and how much the recipient can learn, is limited by the *amount* of information available. The critical factor in this changeover from a purely quantitative theory concerned with the properties of sources and channels, to a semantic one, is "the difference between the study of the conditions of any message whatsoever and the study of the content of *particular* messages" [p.47]. The *average* information carrying capacity of a channel is irrelevant to the understanding of semantic information.

> *Information is a question of both what and* ***how much*** *can be learned from a particular signal and there is simply* ***no limit*** *to what can be learned from a particular signal about another state of affairs ... Channel capacity (as understood in communication theory) represents no limit to what can be learned over a channel (from a specific signal) and therefore represents no*

[3]Note that there are other informal or colloquial usages of the term information which are at variance with the notion described here, but Dretske claims that if you try to carefully tie down what the semantic aspect of the term means, than the central notion must be one of truth.

limit to the amount of information (in the ordinary sense) that can be transmitted [p.51].

Consider the example of an information carrying channel which is capable of carrying only one bit of information. Suppose this channel is used to convey information about the position of a coin on the squares of a chess-board. From an information theoretic point of view the most efficient code is one which, say, indicates whether the coin is on a white square (a binary 1) or on a black square (a binary 0). On average, every time the coin is placed on some random square and the one bit of information sent, the number of possibilities for the position of the coin, available to the recipient of the information, is reduced from 64 options (which needs 6 bits to be fully specified) to 32 options (which needs 5 bits of information to be specified). On average the same amount of information is transmitted each time, and this amount is one bit, so the channel capacity is fully utilized. Suppose on the other hand, by prior arrangement, the recipient knows that if a binary 1 is received that the coin was placed on *one* particular special square, and if a binary 0 is received the coin was placed on some one of the other 63 not-so-special squares. Now, when a 1 is received the recipient knows that the coin is on one particular square and the number of possibilities for the position of the coin is reduced from 64 (presuming they had no idea of its location before the message is sent) to just one possibility, i.e., the position of the coin is fully specified and the equivocation is zero. If r_1 is the receiver symbol corresponding to a binary 1, then the conditional probability $p(s_i/r_1) = 1$ if $i = 1$ (say) but 0 otherwise. So the equivocation for r_1 is given by

$$\begin{aligned} E(r_1) &= \sum_i p(s_i|r_1) \log p(s_i|r_1) \\ &= 0 \end{aligned}$$

In this case *6 bits of information* are sent down the 1 bit channel because

$$I_s(r) = I(s) - equivocation, \quad \text{and} \quad I(s) = 6.$$

In the case that a binary zero is received, the recipient knows that the coin is placed on some one of the 63 squares – that it is not on the special square. The number of possibilities is reduced from 64 to 63 so the conditional probability $p(s_i/r_0) = 1/63$ for $i = 2,...64$ and 0 for $i = 1$. Thus the equivocation for r_0 is given by

$$
\begin{aligned}
E(r_0) &= \sum_i p(s_i|r_0) \log p(s_i|r_0) \\
&= 5.977
\end{aligned}
$$

and consequently only *0.023 bits of information* is received. Of course receiving a binary 1 does not tend to happen very often – on average only once in every 64 random placings of the coin. Most of the time very little information is being transmitted – on average only approximately 0.11 bits – and this code makes very inefficient use of the channel capacity. Nevertheless, this code allows the possibility of fully specifying the position of the coin (albeit not very often), and thus transmitting 6 bits of information. Thus we see from this example, that the amount of information sent in a particular signal can be evaluated, and can be different from the data content of the signal. So the inescapable conclusion is that if we want a proper measure of information content rather than data content, we must concentrate on the amount of information contained in *particular* signals. This is quite different from the usual approach to information theory in the sense of communications, where statistical averages over ensembles of *possible* signals are the prime focus of concern.

7.2.2 The information content of a message – equivocation and truth

The amounts of information contained in particular signals can in fact be evaluated with the aid of ideas involved in the derivation of formulae in communication theory. The notion of the surprisal $I(s_i)$ of a particular event s_i is one of the quantities of interest. Another is the *equivocation* used above. In communication theory, where equivocation is unavoidably present, the *most efficient* codes result from the equivocation being evenly spread over the symbols or possibilities of the code. This is illustrated in Figure 22(a) where the two bits of equivocation is equally spread over the two symbols r_1 and r_2 (2 bits each). On the other hand, the greatest values of surprise occurs when some symbols of the code have no equivocation associated with them, while whatever equivocation exists is spread over the others. This is illustrated in Figure 22(b) where there is no equivocation associated with the event s_1 and information about it received by r_1, but there is substantial equivocation associated with the events s_2-s_8 and the information about them received by r_2. Whether or not a particular signal is equivocal depends on how the possibilities at the source are partitioned relative to the receiver (or the signal itself).

In order to get a definite measure of the information generated by a source or arriving at a receiver from a particular source, we must be able to evaluate the distribution of probabilities associated with the full range of possible outcomes of the source and receiver events. This is possible in the toy scenario envisaged above, or in typical

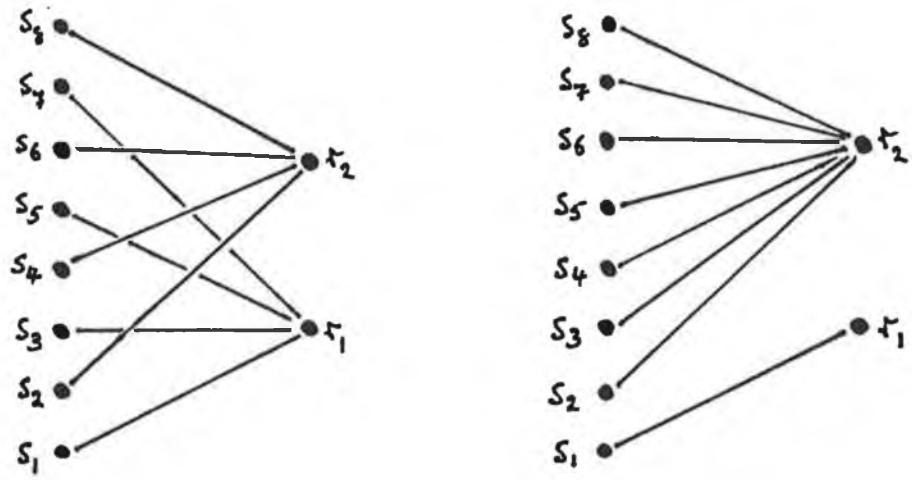


Figure 22. Illustration of the connection between source and receiver events (symbols), in the cases of maximum average information capacity (a), and maximum surprise (b).

communication engineering applications, but in ordinary everyday situations it is not possible. In many circumstances it may not even make sense to ask the question "How much information is being transmitted?", because the range of alternative possible outcomes is not well-defined. It usually is possible nevertheless, to compare the amount of information generated by a particular event, with the amount of information a signal carries about that event, without evaluating either of the absolute amounts of information separately. The crucial factor which needs to be evaluated to make these comparisons is the *equivocation.* If the equivocation at a receiver, of the information generated by an event at some source, is *zero*, then the signal arriving at the receiver carries *as much information* about the source, as is generated by the particular event at the source.

> *The epistemological application of communication theory, and in particular the use of this theory in developing an account of a signal's* ***informational content*** *(the message it bears), does not require that we know whether a subject has received 20, 200 or 2,000 bits of information. All it requires (quantitatively) is that the subject receive* ***as much information*** *as is generated by the state of affairs he knows to obtain [p.55].*

An important point that needs to be stressed is that the conditional probabilities used to define the equivocation between the source and receiver are intended to be *objective* quantities. Whether or not they are known or can even be evaluated is irrelevant to the value of the equivocation and therefore irrelevant to the relationship between the information carried by the signal and that carried by the source. "How much information a message carries is *not* a function of how much information the recipient thinks it carries. It is a function, simply, of the actual possibilities that exist at *s* and the conditional probabilities of these various possibilities after the message has been received" [p.55].

To sum up this idea about the amount of information a signal carries about a source: "if a signal carries the information that *s* is *F* (where *s* is some item at the source), then the *amount* of information that the signal carries about *s* must be equal to the amount of information generated by *s*'s being *F*. If *s*'s being *F* generates 3 bits of information, no signal that carries only 2 bits of information about *s* can possibly carry the information that *s* is *F*" [p.58]. Note that Dretske is speaking here about a particular signal like the signal that is generated when s_1 occurs in Figure 22(b).

The quantitative fact that

(i) a particular signal carries as much information about *s* as would be generated by *s*'s being *F*

is not a sufficient condition to define the content or information carried by a message. A second condition is that

(ii) it must be the case that *s* is *F*.

In other words the message must be a true message. In addition, to ensure that the message is not only carrying enough information, but enough of the right information we need a further condition:

(iii) the quantity of information that the particular signal carries about *s* is that quantity generated by *s*'s being *F*, and not for example, by *s*'s being *G*.

These three conditions collectively are the basis for Dretske's semantic theory of information. They clearly mark out the separate requirements of the message in order that its information content be described. In a more precise reformulation, Dretske also includes the notion of what the receiver might know about the possibilities that exist at the source:

Informational content:

> A signal r carries the information that s is F = The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1).

7.2.3 Analog and digital coding of information

Our interest in Dretske's semantic theory of information is that it allows us to give a precise definition of what is meant by the intuitive concepts of *implicit* and *explicit* information used here. Dretske does this using the terminology of *analog* and *digital* coding in a quite suggestive, if unorthodox way. The usual distinction between analog and digital is intended to distinguish between the notions of a continuous and a discrete representation. Interpreting this distinction in informational terms however, in terms of the way *facts* can be represented, leads to a more general understanding of the concept than that applying just to the representation of physical magnitudes. Thus "a signal carries the information that s is F in *digital* form if and only if the signal carries no additional information about s, no information that is not already nested in s's being F. If the signal *does* carry information about s, information that is *not* nested in s's being F, then I shall say that the signal carries this information in *analog* form" [p.137]. This statement about the relationship between the different forms of information that a signal can carry *about* a source is, I submit, equivalent to the intuitive distinction between implicit and explicit information. It also captures the idea that every signal carries information in both analog and digital form or carries both implicit information and explicit information. The most specific piece of information that the particular signal carries about a source, is the only information about the source which is explicit – the only information which is in digital form. In contrast, when a signal carries information about a source, for example that s is F in analog form, there is always some more specific information about the source contained in the signal than the information that s is F.

Dretske uses the example of the difference between a picture of a cup with coffee in it, and the statement "The cup has coffee in it" to illustrate the distinction between the two forms of information. The statement (in its usual meaning and use) carries information about the cup and its contents in *digital* form. No more information about the situation is available from this message or signal. The picture on the other hand also carries this information, but this is not the most specific information that the picture carries, nor is it the only information. The picture might also carry all sorts of subtle information about the size, shape and texture of the cup, the amount and colour or the coffee and so on, but these are not explicit. What is explicit in this case, what is the most specific information carried by the picture would be the sampled and quantized intensity values of the pixels making up the image (presuming that it is a digitized image in the conventional sense of digital), or the analogue light intensity emitted by the phosphors on a TV screen (in the conventional sense of analogue).

Another example described by Dretske which illustrates an important aspect of the relative transition between analog and digital, is of a digital speed warning system[4]. Consider a digital speed signal in a car, perhaps coming from a computerised engine control system. The digital signal is quantized in steps of one, in the range zero to one hundred. This is a digital representation of the speed in both senses of the word used above: the speed value is the most specific piece of information that this signal carries, and the value is quantized. Now suppose that this signal is further transformed to give the driver an indication of what gear they should be driving in, given the road speed. This transformed signal might consist of a number of discrete ranges:

0-14:	first gear,
15-24:	second gear,
25-49:	third gear,
50-99:	top gear.

Now, in this new (lets call it the "gear") representation, the most specific piece of information is which of the four gears in being indicated. In the previous (lets call it the "speed") representation, this gear information was implicit. Relative to the gear representation, the gear information coded in the speed representation is in analog

[4]An important point to note is that the terms *analog* and *digital*, in the sense that Dretske uses them, are always relative. What might be analog in terms of one source and the information about it, could actually be digital in terms of another related source.

form. The speed information generates 6.64 bits of information, but there is positive equivocation involved in the transformation or classification involved in the speed to gear transition. Because of the different speed ranges classified into the individual gear indications, each of the (four) gear values carries different amounts of information: 2.75 bits, 3.32 bits, 2 bits and 1 bit respectively[5]. The process of transforming information from an analog to a digital form is always one that involves a *loss* of information: "Until information has been lost, or discarded, an information-processing system has failed to treat *different* things as essentially the *same*. It has failed to classify or categorize, failed to generalize, failed to 'recognise' the input as being an instance (token) of a more general type" [p.141]. We show next how this is related to the process of going from "signals to symbols".

7.3 Going Symbolic

Several authors, such as Marr [9], and Wilson *et al* [166, 201,167] have described the process of perception in terms of "going symbolic". In Marr's case this is seen as a once-off transition that takes place as early as possible in the processing stages of visual information

> *Most would agree that an intensity array I(x,y) or even its convolution $\nabla^2 G * I$ is not a very symbolic object. It is a continuous two-dimensional array with few points of manifest interest. Yet by the time we talk about people or cars or fields or trees, we are clearly being very symbolic, and I think again that most would find suggestions of symbols in Hubel and Wiesel's (1962) recordings. Our view is that vision goes symbolic almost immediately, right at the level of zero crossings, and the beauty of this is that the transition from the analogue arraylike representation of the discrete, oriented, sloped zero-crossing segments is probably accomplished without loss of information [9, p.343].*

The careful analysis of information theoretic ideas in this and the previous chapters was carried out so that exactly this type of confusion on the meaning and role of symbols could be eliminated. The most obvious problem with Marr's statement is the notion of a transition to symbols happening *without any loss of information*. By definition a classification, or so-called signal-to-symbol transition must involve positive

[5] A subtle point about this example which is nonetheless important for understanding the nature of perception is the following. Even though the information about what gear the driver should be in, is contained implicitly in the speed value (say 30kph), it is the classification process that makes this explicit and decides what it means. There is nothing intrinsic in the signal "30 km/h", regardless of its representation, which also says "3rd gear". If there was we could make a logical deduction: *if speed is 30 kph, then gear must be third*, but of course we cannot.

equivocation. It must involve some notion of invariance – some notion of *different* input events, states or signals being somehow grouped together, or considered as *alike* in some way, and labelled with a single symbol. In other words, each symbol must represent a whole class of signals – it is a process of generalization in which symbols express invariances or define equivalence relations among the set of signals. That is, according to the definition of what symbols are and what they do, the process of "going symbolic" *necessarily* involves a loss of information. Without a loss of information, we cannot associate two different input signals and call them the same. One possible source of this confusion of the role of symbols is the emphasis on the *referential* idea of symbol without considering fully the information theoretic relationship between signals and symbols.

Cariani describes some more of the salient characteristics of symbols:

> *Symbols are only possible as discrete alternatives implemented through distinguishable signs via material tokens. Without alternatives nothing can be communicated, without discreteness, the alternatives cannot be distinguished [28, p.20],*

but he too makes the second mistake that is inherent in Marr's description of "going symbolic" above. Consider the diagram shown in Figure 23, which illustrates the relationship between the notions of explicate (symbolic) and implicate (non-symbolic) as used by Cariani. In this diagram and in the text, he considers a single level of transition between non-symbolic and symbolic (called measurement) and a single level of transition between symbolic and non-symbolic (called control). Everything in the upper half of the diagram is completely symbolic and involves formal computational processes. As we attempt to show in Dretske's account of the "analogue to digital" transition, and in our own discussion below on implicit and explicit information, the notion of a measurement, or "signal-to-symbol" or "analogue-to-digital" transition is a relative one, involving a generalisation over the input signals, and *necessarily* involving the introduction of equivocation between the input and the output of the process of transition. But there is nothing stopping us from carrying out *further* relative classification processes, introducing *further* amounts of positive equivocation, either on one given signal input which has already been classified, or across several input signals, each of which have already undergone some sort of a generalisation or classification process. The important thing to remember is that despite our appending the abstract notion of a symbol or a label to the output of a classification, the output

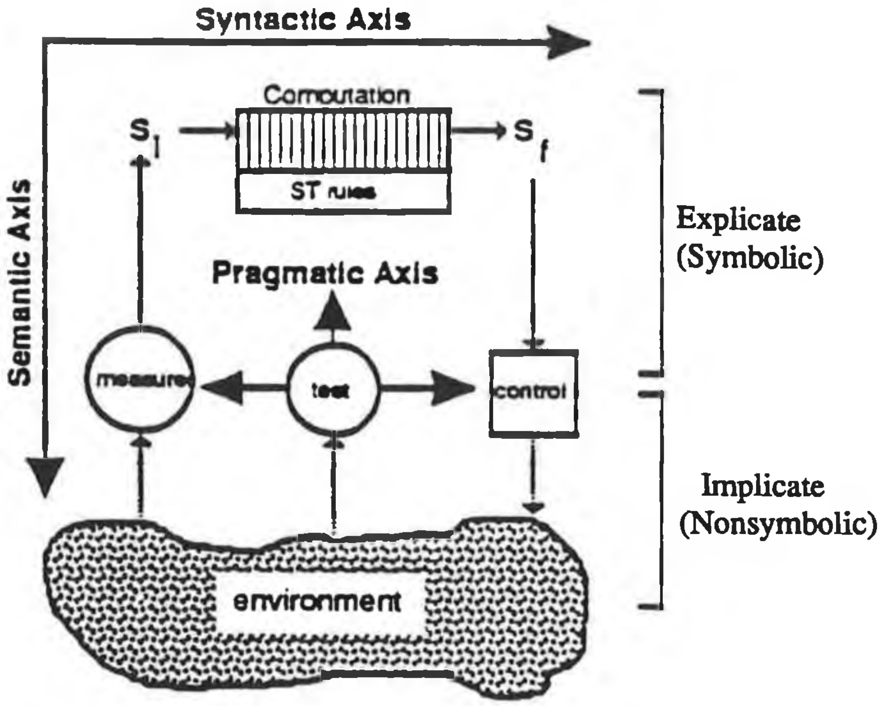


Figure 23. Illustration of the relationship between explicate (symbolic) and implicate (non-symbolic) in Cariani's basic semiotic functionalities of organisms and devices. From [28, p.75].

is just as much a signal as the input[6], albeit a different signal with different properties and carrying less information, but a signal nonetheless[7].

Recall that during our discussion on universals (or concepts or generalization or classes) we described them as things that could be named — that could be represented

---

[6]Cariani uses the term "sign" to denote the physical token embodying or carrying the associated abstract idea of a symbol. The point being made here is that there is no need to distinguish the output "sign" from the input "signal": they are both embodied as some physical value or property. Certainly they are not the *same* signal, but introducing the distinction of a "sign" only serves to hide the relative nature of a generalisation or classification or measurement.

[7]Despite clearly pointing out the *relative* nature of what he calls an "analogue-to-digital" transition or generalization, Dretske actually goes on to take exactly the position as Cariani on the relationship between sensory (analogue) and cognitive (digital). Here we do not pretend to clarify the status of "cognitive" processes, but we do emphasize that there is not one single level of transition between signals and symbols, but rather the possibility of many such transitions.

by, or are equivalent to, *symbols*. The input to a visual system is ambiguous in many ways. It is a flux of signals impinging from many different sources and is capable of carrying to the perceptual system, perturbations correlated with (or carrying information with respect to) the properties of these sources. Much of the information about the external world is *implicitly* contained in the distribution of grey level (or colour if appropriate) intensities across the image (pixels). The intensities are the only quantities that are *explicit* at this stage[8]. Notice that the intensity detected at a particular pixel is one of a mutually exclusive set of possible values -- there is no "vagueness" about the value -- one and only one is selected (deemed to be appropriate). This notion can be extended to any quantity which one wishes to know, or which is required to be measured. The only way we can *know* something is to *decide* that it is one and only one of a set of possible alternatives, i.e. to make its value explicit if a numerical representation is suitable. There can be no overlap, superposition or ambiguity about the answer: a quantity is only known if it has a particular value. This is a straight-forward *classification* process, which is one of the vital steps in classical pattern recognition processes. If the value of a quantity or its classification is ambiguous -- if there are a number of alternative possible classifications, if the classification is anything other than trivial, then the classification process is *inductive*: there is no necessary solution[9]. But furthermore, not only must we *decide* what value to pick if there is any ambiguity, but we even determine what information we are going to extract by deciding what measurement to make. *We* determine what is important in the flux of signals perturbing our perceptual system. We are not simply trying to extract some objective, if ambiguous, information which is the only information that is available.

The information implicitly contained in the intensities of a grey-level image is highly ambiguous. There is a (virtually) infinite number of possible interpretations that could be made with many different variables. Most of these variables we know nothing about

---

[8]Recall the discussion on direct and indirect "detection" vis-a-vis the photoreceptors above -- a *detector* is a device which makes the value of a quantity *explicit*.

[9]This process of deciding on one out of several possible alternatives is also referred to as *measurement*, particularly in the context of quantum mechanics. In substance it is the same as the common-sense notion of measurement: assigning one particular value out of an often continuous range of values, although in the common-sense case, the assignment is trivial (only one value is possible), while in the quantum mechanical case, there is often a probability distribution over the range of possibilities.

– they are combinations of the input data which capture in some way the statistical structure (in the sense of redundancy) contained in the image. Some of these variables do have names and include things like orientation, shading, illumination, texture, depth etc., etc. But the most important thing to remember about the meaning of these variables is that they are not some *ad hoc* "knowledge of the world", which has been acquired by undetermined means, to assist in the process of interpreting or inferring hypotheses from an inadequate and ambiguous image representation. The basis of the meaning of these variables, and therefore of the perceptual decisions involving them, can only be the relationship between this observation and others in space and time (made by us), captured by the statistics of the ensemble. This emphasis on observation and the relationships between observations, as opposed to considerations of objects and their properties, is a point of view introduced to pattern recognition by Watanabe and originally inspired by some of the ideas of Plato[10].

We have seen that the step of going from signals (containing implicit ambiguous information) to symbols (where information is explicit and unambiguous), is vital to the process of perception – without this there can be no perception; no knowledge. It was suggested that this step is somehow related to the redundancy contained in an ensemble of images. We will now try to make these connections somewhat more transparent. One of the connections is particularly simple: if there is implicit information there is ambiguity (except in the trivial case). If we want to make information explicit we should try to get rid of ambiguity – or at least make it as small as possible before the measurement or classification decision. This means making the probability of class choices as *unevenly* distributed as possible, or equivalently, minimizing an entropy function over the set of possibilities for the values of the explicit information.

7.4 Measurement Aspects of Perception

The framework within which perception is discussed here, is based on the assumption that both observer and observed affect each other. This can happen in a number of

[10]For example, Watanabe considers the object-predicate table as simply a set of relationships between different observations: those which we use to "recognise" "objects", and those which we use to "recognise" "properties".

different ways, but the application of propensity is solely on the basis of reason (a) in section 6.5.4 above: the effect of observation on the observed.

Before proceeding it is important to clarify what we mean by observer and observed in this context. The natural assumption is that the observer is the person or machine involved in "looking at" and interpreting the external world which is the observed. This is *not* the case here in this chapter[11]. We use the term "observed" is a sense which is much clearer, if we ignore the distinction that is drawn between input (signals) and output (symbols) of a classification process. Initially the "observed" is the sensory data input to the observing system (the input signal). Subsequent to the act of observation the "observed" is the result of the classification process, which is necessarily a part of observation if information is to be made explicit (the output signal). The observer is not the entire person, but the functional unit which is responsible for taking some small fraction of the sensory data (or processed versions of it) as input, and producing a classification of that input into exactly one of several mutually exclusive classes, thereby introducing equivocation to that signal. In order to distract the reader from the usual engineering linear input/output system conceptualization, it might be useful to make the following analogy: the observing mechanism is like a phase-locked loop that can "lock" onto one of a number of discrete frequencies.[12] If the circuit is excited with energy at a frequency other than one of the selected ones, it might use that energy to excite one of its preferred modes — perhaps the one nearest to the applied frequency. If energy is applied spread over a wide range of frequencies, it might be difficult to know which of the discrete modes will win-out and be selected, but one and *only one* will. It will *never* operate at any frequency other than one of its preferred modes.

In line with the ideas proposed by Wilson *et al* [166, 201,167] we can consider the sensory data as defining a probability distribution over the set of possible outcomes

---

[11]In discussions in other parts of this dissertation, particularly those associated with the enactive approach to perception and cognition, the term *observer* is used in something like the conventional sense specifically being excluded in this section, although there it has the additional connotation of someone separate from the machine or organism being focussed on. In this chapter the specific sense in which the term *observer* is being used is a sense akin to the notion of a quantum mechanical measurement apparatus.

[12]No real phase-locked loop might act in the way imagined here but we will consider as a "thought" system in the spirit of the "thought experiments" in theoretical physics.

for a particular primitive measurement or observation process. Then, consonant with Watanabe's notion of the "preparation" of a quantum mechanical state, we define the A-test (the test of A-ness) to be a *random selection of one of the possibilities* allowed by the polychotomy defined by the predicate A, where the *likelihood of a particular selection is given by the probability distribution*. That is, the A-test is the generation of a sample of a discrete random variable with the appropriate probability distribution. If our N sensory data are represented by an N-tuple of complex numbers in a linear complex vector space $\mathbf{C}^N$, then an observation will correspond to the application to an initial vector v a Hermitian operator A to yield a final vector u:

$$u = Av.$$

In general the vector u will have several non-zero coefficients corresponding to the likelihood of different possible classifications in the A-test. Remember, just as in quantum mechanics the only possible outcomes of the A-test are those represented by the eigenvalues of the A operator. In other words, to make an A-test here means deciding between one of a number of possible outcomes[13]. No result is allowed which is not one of these outcomes. No result is allowed which involves thinking that two mutually exclusive and therefore contradictory outcomes can simultaneously be the case. The predicate A defines a number of possible mutually exclusive outcomes, which in the geometric vector space representation, are represented by the eigenvalues of the corresponding measurement operator A. The result of applying A to a vector is another vector which *must* be one of the eigenvalues of A. A result of any other vector is simply not valid.

The non-zero co-efficients of the vector u which result from the application of the operator A to the input vector v are best thought of as the values for the projection of v onto the corresponding eigenvectors of A. Then, as described by Wilson and Granlund [167] we need a decision rule for assigning probabilities to the eigenvectors of A (which are the only possible outcomes of the measurement of A), considered as final vectors. If the vectors $\varphi^A{}_n$, $n = 0,\ldots, N-1$ are the eigenvectors of A then we need a rule which gives the probability $p_n(A,v)$ that $\varphi^A{}_n$ is a final vector, given that the initial vector is v. The natural choice for a general vector uses the co-efficients of

[13]In the case of a measurement of position the decision is between one of an uncountably infinite number of possibilities; the probability distribution becomes a continuous distribution; and the possibilities are represented by an uncountable infinity of position eigenfunctions in an infinite dimension Hilbert space. The underlying ideas are not however changed in substance.

v expressed as a linear combination of the φ^A_n's (u is the representation of v is the co-ordinate system with the eigenvectors of A as basis vectors).

$$v = \sum_{n=0}^{N-1} v_n \, \varphi^A_n \qquad p_n = \frac{|v_n|^2}{\sum_k |v_k|^2}$$

Note that the output of the measurement of A is not a simple expansion of v in terms of the eigenvectors of A. Rather the process described here is one of inference or decision: "pick the eigenvector φ^A_n which best represents v and if you cannot decide, toss a (n-sided) die to help you" [167].

7.5 Implicit and Explicit Information

We are now in a position to make clear the difference between implicit information and explicit information and the role of primitive perceptual observations. Primitive perception involves going from the particular to the general. It is an inductive process which involves a non-necessary extra-evidential decision about a general concept (a predicate) applied to a particular object (our input data). Implicit information means nothing on its own – it does not exist, except as an interpretation or concept assigned or applied to the input data by an external agency – in this case the observing mechanism. Implicit information is not knowable. The same data can be assigned infinitely many interpretations (infinitely many different measurements of concepts or predicates can be applied) and so the data can be *thought of* as containing all these different types of information implicitly. The implicit information attributed to the input data set by virtue of a decision to make a particular observation on this data set (to measure or classify as per a particular predicate) is usually ambiguous. The data, interpreted as containing a particular type of implicit information, defines a probability distribution over the set of possible outcomes of a measurement defined by this predicate.

Explicit information is the decision that a particular case (or possibility) of the set of possible outcomes defined by this concept (predicate), is the one that was implicitly contained in the input signal. But there usually is no necessary result for this decision – any one of the possibilities with non-zero probability could have been (and if the measurement were repeated on the same data, sooner or later would have been) selected. Note that in explicitly deciding that one particular possibility (whatever its probability defined by the data set) is implicitly described by the input data, we are loosing implicit information about other possible, but incompatible interpretations that could be made. If we wish to simultaneously know the result of two different observations or measurements, the input for the second must be the output of the first. We can do two different measurements of two different predicates or concepts on the original data (like to measure the position of edges and measure the class, or type of edges) but the two explicit results are not simultaneously true. The very statement that we would like to make: that an edge of a certain class lies at a certain position, is exactly the statement or conjunction that we are not allowed to make when the measurements are incompatible. At least we are not allowed to make this statement with a joint accuracy or specificity in the pair of observables, which is greater than a certain amount defined by an uncertainty relation [166, 201]. The information lost when one possibility defined by the first predicate (the position) is selected, is exactly the information needed to decide between members of the second set of possibilities (the edge type or class). The non-distributive or quantum logic involved in primitive perceptual observations enters the system for the following reason. If there is any doubt, any ambiguity about which class a signal falls into (or the value of a measurement on input data), and the signal is subsequently classified into one particular class defined by the concept or generalization, then the propensity theory (for which the distribute law does not hold), immediately applies. Deciding that some signal *does* belong to a particular class, changes that signal if there was any ambiguity about which class it belonged to before the observation.

Implicit information is an interpretation of data which is ambiguous. Within this interpretation, implicit information defines a probability distribution over the set of unambiguous results. Only if a decision is made about which class or result is selected, or what state the perceiving systems enters as a result of the input data, does the information in this interpretation become explicit. So this particular inductive ambiguity – which class or result is suitable to describe the input – is overcome by using the available data to determine a probability distribution and randomly picking a class, consistent with this probability distribution. The notion that a random process is involved in deciding what particular class or measurement value is appropriate when making a primitive observation, sits a bit uncomfortably at first. One alternative is to pick the most probable result each time, but on average this rule fails to capture the essence of how the input data implicitly weighs the various possibilities. Wilson and Granlund give a thorough analysis of why the rule for deciding between different possible alternatives in a classification process is appropriate [167].

7.6 Adaptation or Metadynamics

It was mentioned above that as well as the observing system affecting the observed data, at least as far as subsequent observations are concerned, the observed data has an effect on the observing system – the observer. While the observer can have a drastic effect on each individual set of input data, it is assumed that the reciprocal process of the observer adapting to the data involves such small changes that it is only over the ensemble of data (or a representatively large sample from it) that the development of the observer's response can be seen. Nevertheless, it is now becoming clear that statistical properties of the input data are appropriate to control the development process of perceptual mechanisms in order to achieve some result which seems to be desirable for a perceptual mechanism. We describe the work done by Linsker above which shows how a simple Hebbian-type of adaptation in a multi-input adaptive linear system has the effect of maximizing the amount of Shannon-information flowing through the system. Recent findings by Barlow and Foldiak show that neurons adapt to decorrelate their response to give the sort of effects described above [17].

A much broader view of metadynamical procedures for changing the dynamics of the system has been taken in the area of artificial life (Alife). Varela has categorized at least three different metadynamical strategies [53]:

(i) **neuronal strategies**: the number and types of nodes in a network are fixed throughout the period of adaptation, the connections vary in number and/or strength according to some rules (e.g. Hebb rule);

(ii) **genetic strategies**: the emphasis is on the application of rules for replacing and updating nodes (there may be no interaction between nodes within the adapting system) [205,206,207];

(iii) **immunological strategies**: the connections between nodes are not modified *per se*, but the list of active or participating agents (nodes) changes continuously – not as a recombination of old nodes, but in the form of new recuits [208].

Consider a dynamical description of an autonomous system's closure giving rise, not to a unique solution, but to an ensemble of possible solutions or trajectories for the system. In order that the system remain viable the system must pick a trajectory so as not to depart from the domain of constraints which guarantee the systems continuity – the viability subspace [209]. At any moment the system must guess an appropriate set of solutions by eliminating all the others. This process is basically what Peirce refers to as abduction:

> *the capacity to guess the hypothesis with which experience must be confronted, leaving aside the vast majority of possible hypotheses without examination [53].*

The metadynamical strategies above are intended to ensure that the operational closure of the system is respected, not only in the state space, but also in this continual change in the defining dynamics so that the chances of remaining in the domain of viability are maximized.

7.7 Symbolic Descriptions

Before leaving this topic of the non-distributive logic that underlies perception, it is important to point out the relationships *between* the primitive perceptual (classification) processes described here for making information explicit. According to Wilson and Knutsson [166, 201], the primary goal of visual processing is to obtain invariant (symbolic) descriptions from the signals constituting an image. By a symbolic description, these authors mean anything from a low-level description in terms of primitives such as line and edge elements, to a description in terms of solid objects occupying a definite 3-dimensional volume. Here we make no such attempt to define what the symbols represent or the observations measure. Rather, we believe, like Barlow, Field and Linsker, that the perceptual mechanisms adapt to the statistical properties of the input data, usually during a certain critical or sensitive phase of development, in order to optimize some function like the ability to extract (make explicit) information; to maximize discrimination, etc. We can probably make *a posteriori* guesses about the functions of particular perceptual sub-systems in terms of our cognitive concepts of the real world like edges, surfaces or volumes, but we must be careful for two reasons:

(i) Such interpretations are notorious for being wide of the mark. For example, work on artificial neural networks has shown that cells in the cortex which were interpreted as being edge and corner detectors using the classical bar and edge stimuli of neurophysiology, may be better interpreted as estimating 3-d shape from shading, and measuring curvature, respectively [210,211,212].

(ii) These interpretations also give no clue about why the system developed in the particular way it did; what the mechanism of development was, or what variable or function was being optimized. As Laughlin [15] and Barlow [16] both claimed, our cognitive judgements about the perceptual importance of things such as edges and surfaces etc., are probably more *caused by* our perceptual systems than *explaining* them.

Wilson and Knutsson's treatment of the mathematics of the extraction of symbolic descriptions of the world is independent of any interpretation of what these symbols represent and so is consistent with our treatment here.

> *The essential property of symbols is not, therefore, their position in the hierarchy of image description, but that each symbol represents a whole class of symbols ... it follows that symbols express certain invariances among the totality of signals ... Another way of expressing this idea is to say that symbols carry a notion of identity. Two signals are regarded as identical if they give rise to the same symbol ... symbols define an equivalence relation among signals [166, 201].*

So, as well as the ambiguity that exists over which symbol best represents a particular signal, one symbol can represent many different signals. Again, we consider our input data or signal as an element of an N-dimensional complex linear vector space $\mathbf{C}^N$; (it could be a sampled 1-dimensional signal). The condition that probabilities remain unchanged despite changes in luminance for constant contrast is important. It means that all vectors belonging to the same 1-dimensional subspace (that have the same direction but different lengths) are equivalent. In general, the equivalence classes for a complete set of symbols correspond to the invariant subspaces of some suitable transformation. The elements of the invariant subspaces are the eigenvectors of the transformation and for any linear transformation there is a unique decomposition of the vector space into irreducible subspaces. Each symbol is associated with a particular eigenvector and so for an N-dimensional space the symbols are organized into groups of N which are associated with a particular transformation or operator. Different transformations, if they do not commute, give different, incompatible sets of eigenvectors and therefore symbols. Wilson and Knutsson then go on to show that the eigenfunctions of the translation operator (the class-defining symbols which must be invariant to translation) are the complex exponentials, and the expansion of the input signal in terms of these symbols is equivalent to the Fourier transformation. These Fourier co-efficients are the probability distribution over the set of symbols discussed above. The Fourier symbols are one particular way of interpreting the given data vector. In the propensity theory there is an f-function for each particular Fourier symbol measuring the degree of certitude that our input vector can be represented by that symbol alone. The Fourier expansion still implicitly contains all the information in the original data vector but it contains no explicit information – we do not know *which* of the mutually exclusive possibilities (the complex exponentials) should be selected as the one to represent explicitly the implicit information contained in the original data. As soon as the Fourier information is made explicit and one particular Fourier symbol (complex exponential) is selected, other implicit information is lost. The data can no longer be reconstructed. Thus in this case, observation – the inductive decision that a particular symbol is suitable for representing input data – means that some information is made explicit at the expense of an overall loss in implicit information that would allow the determination of other symbol interpretations. This selection of a particular Fourier symbol is similar to the attempt to decide that a particular class of edge exists in an image. In theory there is no limit to the amount of spread that must be allowed to determine the class of edge. But this means that all information about the position of the edge is lost. Position is coded by a different, and incompatible set of symbols.

The position symbols correspond to the subspaces that are invariant when an operator or transformation which switches between classes is applied, i.e. they are invariant to "translation" in the Fourier domain, or equivalently, they are simply invariant to classification. The eigenfunctions are the set of delta functions on $\mathbf{C}^N$ – one for each component of the N-tuple representing each vector. The probability distribution over the position symbols for a particular vector is just the set of components in the N-tuple representation of the vector, i.e., the original data. The making explicit of the implicit positional information in the input data corresponds to selecting one of the positions in the data vector. All the implicit information required to subsequently decide that a particular class of edge, say, exists in this data is then lost. These two systems of representation are fundamentally incompatible in the sense that the class symbols (complex exponentials) carry no positional information and the position symbols (impulse functions) carry no class information. Only limited resolution can be simultaneously achieved in both of these extreme representations and the trade off in simultaneous resolution is described by the Gabor/Daugman uncertainty relations. The symbol set which achieves the limit described by the uncertainty relations are the logons or elementary functions described by Gabor (GEFs). Note that, any particular set of N GEFs (for vectors in $\mathbf{C}^N$),[14] forming a complete set of symbols[15] does not capture all the information implicit in the input data. Parallel sets of primitive observations need to be simultaneously carried out on the same input data. If these parallel "channels" suitably represent different trade-offs in simultaneous localization in the two incompatible domains, the maximal amount of information can be extracted (made explicit) from the input data. (See Field [71], Daugman [134] or Porat *et al* [132] for different ways of tiling the information diagram).

Grossberg and his colleagues have also used similar arguments about fundamental "minimum uncertainties" inherent in visual perception, to motivate a distinction between two very different types of processes, which are both concerned with "edges" or contours in an image [203]. He describes these processes as *boundary contour* (BC) and *feature contour* processes. The reason for concern with contours in the first place is the need to "discount the illuminant". Only at the image projection of scene contours, where a transition in reflection but not in illumination takes place, is there reliable information available about the reflectance. However this information cannot be made explicit in a single process because

---

[14]Strictly speaking the GEFs are only defined in the continuous domain so sampled versions need to be used in the example given here, but we do not want to get lost in the nitty-gritty.

[15]The set of symbols are not complete in the sense that they cover all of Gabor's information diagram. They are complete in the sense of a complete set of N-eigenvectors in an N-dimensional space. The set of symbols corresponding to the eigenfunctions of a particular transformation are equivalent to one of Field's channels.

> *... there exist fundamental limitations on the visual measurement process – that is, uncertainty principles are just as important in vision as in quantum mechanics. For example the computational demands placed on a system that is designed to detect invariant colors are, in many respects, complementary to the demands placed on a system that is designed to detect invariant boundary structures [213].*

This is why the BC systems and the FC systems must process incoming signals in *parallel.* Now Grossberg does not explicitly couch the argument in terms of incompatible sets of symbols or subspace inference, but the effects are the same. Certain classifications or labellings are not simultaneously decidable, so either a level of minimum uncertainty is accepted in a single processing pathway, or two or more parallel pathways need to be involved in the classification processes representing different compromises in the tradeoff of incompatible variables or concepts.

7.8 Perception as a Grouping Process

The two complementary aspects of the term "pattern recognition" discussed above reflect the fact that in order for recognition to take place grouping processes are required

(i) to allow different objects to be collectively represented by one symbol (generalization) and

(ii) to allow a single object to be described as a collection of parts (gestalt instantiation).

These two different grouping processes are different facets of a single function – they can be separately described but any pattern recognition or perceptual processes other than the primitive observations discussed above involves them bound in inseparable dual roles. One of the central theses of this report is that general perceptual sensory processes can be described in terms of functional components which are normally associated with pattern recognition. Thus the basic process of knowledge apprehension by which explicit interpretations of sensory data are made, based on past experience, is an inductive inferential classification process. This process, which we have referred to as a *primitive observation* or measurement, is based solely on the data supplied to it and the statistical properties of the ensemble from which it is drawn. It is not directly affected by other primitive observations in the hierarchy of perceptual mechanisms, except in so far as some of these provide its input. The hierarchy is arranged into various parallel and serial pathways to cope with the conflicting (incompatible) requirements of different interpretations (e.g. form vs motion, BC vs FC), of information implicit in the constant stream of sensory data. The primitive observations however, are only one side of the perceptual duality. The perception of real extended patterns and objects demands that many different primitive observations at different levels of abstraction be somehow associated with each other. Recent results in neurophysiology discussed below, indicate that this linking process may be implemented by the *synchronized activity of neural circuits.* We also discuss below the mechanisms by which the neural connections required to subserve both the process of primitive observation and the association or linking circuits, are set up. Before this however, there are two other aspects of perception that need to be examined – attention, and bias or value-weighting.

7.9 Axiology and Value-Weighting

We have already discussed the fact that pattern recognition is a non-necessary, non-logical, extra-evidential process. The theorem of ugly duckling makes it clear, that without some "preferential ponderation" of variables, paradigm-based pattern recognition is impossible. Watanabe's "axiological overcoming" of this impasse involved the introduction of an "extra-evidential or extra-logical element" to the basic process. In other words depending on the usefulness of a classification, some predicates should become more "important" than others. "Useful" is used here in the sense, not of "any ultimate value, but various instrumental values towards more fundamental ends" [19, p.84]. Some of these instrumental values are clearly innate:

> *the sensory organs have developed themselves in such a way that the more important predicates are directly observed, whereby importance is mainly dictated by the value of survival. To individual animals, these predicates are ... innate ideas, but they are without doubt results of the long experience through evolution [19, p.85].*

In the case of discrete predicates, the value of a predicate can be recognised by assigning a weight to it. The values of the set of weights can change depending on the usefulness of the classification which they allow. We can only see similarity and in fact, can only see at all, because of the "uneven" emphasis on the empirical data from the external world. One useful way of measuring "unevenness" is with the entropy function and a useful way of assigning weights to variables is as a function of their variance. These two ideas, coupled with the minimization of the entropy (which is equivalent to the "principal of simplicity") defined over the variance weightings, and the elimination of derived predicates (variables) with small weighting (dimensionality reduction), form the complete mathematical basis for the principal axis transform and the Karhunen-Loève transform. It is important to remember that these extra-logical extra-evidential considerations are heuristic guiding principles. They do not guarantee a correct solution to the problem of overcoming inductive ambiguity, though often prove to be successful or useful in particular cases. In the next chapter we describe how dimensionality reduction, based on a modified version of the KLT (inspired by a Platonic view on universals), can be used as a preprocessing stage before classification.

The overall impression we get from our considerations of the problems of biological vision is that sensory perception is not a passive and unbiased transmission of physical stimuli. Rather it is an *active* selective formation of *valuable* information. This idea of "active perception" involving interaction with and exploration of the external world, rather than passive forbearance with restricted input has recently received some attention in computer vision. Normally, active perception implies an active movement for the sensing organ (eye, camera) depending on what has just been seen. A typical example is moving ones head to introduce depth parallax, or looking at something "from different angles" if a single view is not sufficient for purposes. These ideas can be extended to purely internal processes involving active selection of data or processing function within the perceptual apparatus. This leads to a rejection of the notion of a passive acceptance of a real world which imposes its reality on perception without any room for individual interpretation. Each individual organism actively *constructs* its own perception based on statistical properties of real world images, but also on value-weighting (which depends on usefulness, emotions, attention etc.) and on random primitive classification events. Note that the gross structure of the brain and its perceptual subsystems seems to be largely genetically predetermined, but with the exact detail of the wiring depending on visual activity – some of it noise generated in the retina prenatally, some of it as a direct result of visual experience.

7.10 Attention

A weak form of active perception involves the notion of attention. Some work has been carried out in psychophysics and neurophysiology which demonstrates quantitative and some qualitative differences in function depending on attention. For example, the pattern of saccadic movements that people make in viewing a scene has been shown to be influenced both by the scene, and by what questions people are asked about the scene [107]. The phenomenon of attention is closely related to gestalt organization. There have been suggestions that the lateral geniculate nucleus (LGN), a relay station on the path between the retina and visual cortex, which is embedded in a part of the brain shown to be connected with emotion, may play a role in attention [37, 40]. While there have been some tantalizing clues about the nature of attentional mechanisms and their role in perception little concrete can be said at this stage other than that like the gestalt organizational properties of perception, attention and its switching may play an important role in the nervous system.

7.11 Cortical Function

One of the most original voices in the debate on visual perception in recent times, has been Horace Barlow. By simply considering the apparent constraints imposed by physical and biochemical processes on the architecture and processing capability of the eye and brain he has drawn some remarkable conclusions about the nature of visual perception [16]. The quality of human performance is high in comparison with purely statistical inferences [157], which means that the information received by the eye about the external world is effectively utilized. This is despite very strong limitations on connectivity within the cortex. The vast majority of the brain's volume is taken up with connections, rather than the directly active processing elements, (which are the synapses and neural cell bodies). The system seems to be limited, not by the number of processors, but by the difficulty of interconnecting them. In any part of the cortex, connectivity between nearby points is very high, up to distances of about 1mm. The remaining connections over longer distances tend to be relatively sparse (in comparison with the amount of very local connectivity), but nevertheless systematic projections, between different regions of cortex. Given that much of perception involves using information from widely separated parts of the visual field, the question of how the relevant information is brought together must be considered. Barlow proposes three stages:

(i) the improvement of the cortical map in the primary visual cortex by processing similar to spatial and temporal integration,

(ii) the detection of "linking features" and,

(iii) the concentration of this information by non-topographical mapping into adjacent visual areas of the cortex.

The spatial acuity of the visual system is well known, as is its hyperacuity abilities in particular situations. For example, a vernier displacement can be seen by the eye/brain down to a resolution which is an order of magnitude better than what would be expected from the sample spacing and modulation transfer function. Barlow argues that the very large number of primary cortical cells per retinal sampling point is quite sufficient to allow this type of hyperacuity with appropriate computation. Less well known are the temporal characteristics of the eye/brain system. The "noisy" character of individual cortical neural activity means that in many neurophysiological experiments, response activity is averaged over 10-15 milliseconds before results are analyzed. Psychophysical results which show that temporal frequencies higher than about 8 hz are very difficult to see, are in complete agreement with our everyday experience with television and cinema which have full stationary picture update rates of 25 per second and field rates of 50 per second. Recent work by Burr and Ross [214] has demonstrated that the extraction of motion information involves integration over times of the order of 100ms. This is not simply a temporal integration process at a fixed position in the visual field. It is a spatiotemporal integration "following the motion" and has been shown to yield a temporal hyperacuity involving the utilization of timing information which is accurate to the order of 150μs.[16]

Another interesting example is the Pulfrich effect where an illusion of movement in depth can be created by dimming the light reaching one eye (thereby increasing the latency) for an object moving across the visual field. It has been shown that time and position are completely interchangeable in giving the cue to depth. Barlow suggests that the systems good performance for moving objects may be achieved by arranging that the neural signals corresponding to the targets, should pass through a screen of spatial and temporal inhibition.

A system capable of such spatial and temporal accuracy must surely need extreme precision in its connections and their strength. It is believed that chemical gradients,

---

[16]Barlow points out that this value of 150μs should be compared with the latency of neural activity, which is about 100 times longer.

presumably genetically specified, are responsible for guiding fibres to their target areas and positioning the terminals across the target area [215,216,217]. However, the actual synaptic connections, their number and efficiency, are believed to be determined by processes which depend on neural activity of both the projecting and target cells. Much of the projection from the LGN and retina to the visual cortex is already in place before a baby monkey is born. The work described above by Linsker [18] on artificial neural nets and Mastronarde [218] on the correlated firing of retinal ganglion cells has shown that random fluctuations generated in the retina may be responsible for the activity-dependent pre-natal mappings which are constructed in the early visual pathways. Prenatal interruption of neural activity in the retinal projections to the LGN and cortex have shown similar disruptive effects to the deprivation of visual experience during the post-natal critical period.

Gestalt organization was mentioned above as one of the two fundamental aspects of pattern recognition, and possibly of perception. Barlow suggests that the keys to collecting together the information for detecting global properties in spite of the very localized connections in any particular part of the cortex, are what he calls "linking features", and also non-topographical maps[17] between different regions or patches of cortex. Barlow defines linking features as

> *those locally detectable qualities of a portion of the visual scene that in Gestalt terms, cause segregation, or separation of figure from ground. Colour, texture, disparity, direction and velocity of motion, and orientation are examples[18] [16].*

The localized ($\approx$ lmm) connectivity throughout the cortex means that in order for information to be allowed to interact, it must be "brought together". Thus as we have quoted Barlow as saying already, the important thing about any patch of cortex is what information is "brought together" there – for unless cells whose activity represent the information in question are within a small distance of each other they (usually) cannot

---

[17]Topographical maps are maps in which neighbourhood relations are preserved. In other words points are mapped near together, which have similar values of some spatial parameters. In non-topographical maps, points which have similar values of some other parameter such as colour or orientation or motion are mapped nearby in the target region.

[18]It seems [13] that the magno system which comprises only a fraction of the visual processing machinery, (even though it is evolutionally older than the more extensive parvo system) may be responsible for the familiar grouping or gestalt effects associated with perception as well as aspects of depth and motion processing. In short it is principally concerned with the "where" part of the "what versus where" dichotomy identified by Marr [9], i.e. it is responsible with the perception of spatial structure. The magno system is apparently "colour-blind" and at equiluminance, many of its functions cease, with startling effects.

interact. The linking features are detected in V1 (area 17) according to Barlow by a "local analysis of the topographically organised reconstructed image". Projections to other regions of the cortex subsequently redistribute this information for further processing.

Much less is known about the response properties of cells in these extra-striate regions, V2-V5 and the projections to them from the retina and V1. There does seem to be a progressive functional segregation into at least three fairly identifiable parallel pathways for separately processing colour (blob pathway), high-resolution static form (parvo-interblob), and movement and stereo depth (magno pathway) [13]. These pathways are not isolated from each other – there are cross projections and cells often seem to retain some responsiveness to stimuli not nominally coded for in their pathway. Higher cortical areas such as V4 and MT (middle temporal area) do have cells which are fairly specialist in their tuning, but it is not known if they are organized in the way the LGN, V1 and V2 seem to be.

And for the next step – Barlow has his own ideas:

> *what is done with the information that there is a region of the visual field with some common direction of motion, or that collinear orientation detectors are being activated? I like the idea that this information is signalled back to the reconstruction in 17, enabling the area that has the common characteristic to be "flashed" or "cross-hatched" in some way, though I am not sure that this idea would be so appealing if we had successfully banished the homunculus from that area [16].*

This suggestion by Barlow is particularly interesting as it seems to be the first time that anyone has seriously suggested a role in perception for the very extensive back projections from the "higher" visual areas to V1. It also provided one of the impetuses for the theory of perception presented here.

As mentioned above there have been several attempts, which are widely known, to elaborate a theory of perception, i.e., how all the various components or visual properties into which the visual system seems to decompose its input, are re-integrated into a single percept. Hubel and Wiesel proposed a system of hierarchial feature detectors of increasing specificity. An alternative suggestions at about the same time was in terms of Fourier analysis. Marr combined ideas from both, justified by his computational theoretic approach, to describe a succession of filtering processes between well-defined representations which culminated in an object-centred 3-dimensional description. The logical conclusion of the hierarchical feature detector theory is the idea that somewhere in a "higher" cortical area, there could be a "grandmother" cell, with obvious response properties. This idea, which was probably just a "straw-man" originally anyway, is demolished by the inevitable combinatorial explosion.

Recently an alternative notion has begun gaining ground [219]. This is the idea that the representations in the brain of various visual properties of objects in the world are only *transiently* combined, rather than in fixed receptive fields. This is done "in some way that makes the conjoint output of different property-specific detectors available to the mechanisms for perception or action. A possible mechanism for the "transient combination" may be illustrated by the result that "neurons in the visual cortex activated by the same object in the world, tend to discharge rhythmically and in unison. (This gives a whole new meaning to Gibsons notion of the visual system "resonating" to "invariants" in the external world).

One of the reasons why these correlations had not been reported before was the difficulty of recording the activity of one neuron, let alone two or more. With the advent of procedures for recording and cross-correlating data from several individual cells, a whole new vista has been opened up on the brain. Visual stimulation seems to cause many neurons in the visual cortex to fire rhythmically at 40-50Hz or 40-80Hz [220] and the oscillation seems to originate in the cortex. Oscillations in the range 20-80 hz are usually called γ-waves. The oscillations are usually in phase, when recorded from neurons with overlapping receptive fields, irrespective of selectivity for stimulus orientation. For cells more than 2mm apart, (no receptive field overlap), oscillations are only in phase for neurons with the same orientation tuning. For very distant pairs of neurons (7mm; same orientation specificity) the cells' activities were strongly correlated when a single long bar cut both receptive fields. Two shorter bars which did not bridge the gap between the receptive fields did not produce correlated outputs. The correlated patterns of activity seem to encode a global property of the stimulus, in this case whether or not the stimulus is a single contiguous object. The power spectra for the oscillations seem to be broadly distributed indicating a chaotic rather than sinusoidal source [220].

Over the last 100 years, as more and more of the cortex has been identified as being directly or indirectly related to some or other sensory or motor activity, the "uncharted" areas of cortex for which no such link had been established (often called the association cortex), have been steadily decreasing. Barlow's approach with linking features and non-topographic maps puts the cortex into a completely new light. The fact that some areas of cortex contain well-organised neighbourhood-preserving (topographic) maps of the sensory surfaces (retina, cochlea, skin, etc.) is not just so that they can carry a "representation" of the sensory surface. Rather, the information contained in a visual data stream has a natural representation in terms of a large number of parameters, including space, time, colour, motion, orientation, binocular disparity, etc. The topographic maps onto the primary sensory areas of the cortex are simply those maps where the spatial parameters are made explicit and spatial data is brought within the processing range of the cortical machinery. In this way local properties of the visual data can be detected or processed. Local interactions in the formation of these topographic maps can and do give rise to more global ordering of sensory parameters which cannot be immediately mapped within the cortical processing range. Thus we see in the topographically organised primary visual cortex, that parameters such as ocular dominance and orientation selectivity are mapped on the primary cortex in an organized way. But interactions explicitly involving this information cannot be fully implemented at this stage because they are out of range. Subsequent, non-topographic maps to so-called secondary sensory areas on the cortex seem to make these parameters – colour, motion, stereo – more explicit and the spatial parameters less so. This is exactly the type of mapping needed to detect and process more global properties of the data which depend on data which is "nearby" in some sensory parameter "space", and not necessarily nearby or local in physical space. This also helps to explain why there are projections from the LGN (carrying data from the retina) directly to these secondary areas: in these cases the spatial parameters is not the one which dominates in the initial projection but some other parameter associated with this cortical region.

> *In this way the whole cortex acquires its unity again, for it all becomes association area, and the primary projection areas with good topographic maps are simply regions specializing in the detection of local associations ... the natural question to ask about a particular cortical locus becomes "what types of information are brought together here?" rather than "what is represented here? [16]*

This quote from Barlow sums up what we believe is the appropriate way of trying to understand the visual cortex in biology, and in trying to implement artificial systems with visual perceptual components. This idea of the whole cortex as association cortex, coupled with a mechanism for transiently making explicit the associations for a particular stimulus (coherent oscillations), led us to formulate the viewpoint on perception which is presented here. It is probably best not described as a full "theory" of perception as there are several important gaps. It is so to speak an initial position – a theoretical framework – from which a full theory of perception might be formulated.

We believe that there is no hierarchy of visual cognition where features detected in one area are fed onto another area for higher-level processing, and subsequently to even higher areas culminating finally in some cognitive centre which is the seat of consciousness. Rather, we suggest that cognition is the totality of synchronized activity at all levels in all cortical structures at a given time. If some things like edges are detected early on in the processing hierarchy, they are not passed on as a *fait accompli* for cognition elsewhere. This is the only place that edges are detected, or used, or *perceived.* In this way, different facets of perception and consciousness are simultaneously supported in the appropriate processing structures. In the context of Marr's theory of perception, this idea would be like saying that Marr's primal sketch is just as active a part (or fact) of perception as the 2½d or 3-D sketches – not that they are earlier stages which feed their output to higher centres for final representation, recognition or memory storage. There is a hierarchy of processing in the sense that some later processing normally depends on the *activation* of lower-level processing but we do not believe that the information processed or detected at the lower-levels is channelled to the higher centres.

Hubel [61] asks the question of "how all the information [processed by the cortex] is finally assembled, say for perceiving a bouncing red ball?. It must be assembled somewhere, if only at the motor nerves that subserve the action of catching. Where it's assembled, and how, we have no idea". Our light-hearted reply to this question is, "For what purpose would the motor system need to know that the ball is red?" The motor system receives the appropriate abstract information to carry out its function. No, we suggest that the perception of a red bounding ball "out-there" in the external world depends on the coherent activity in all the various visual cortical regions, but crucially, on the activity in V1, the direct topographic link with the external world. If all of the other regions are active but V1 is silent we may be dreaming or imagining but we are not "seeing". If V1 is active because of the action of drugs or whatever, not input from the retina then we *are* "seeing" – but not the real world, we are seeing hallucinations. Recent work on neural correlates of perception seems to confirm that V1 is not active during imagery but that other visual cortical regions are [221]. To summarize, it is fairly clear that each visual area brings together particular types of visual information, and a particular stimulus with a range of characterising attributes simultaneously arouses activity in parts of many of the appropriate visual areas. We suggest that the collective synchronized excitation of cells in all of these particular regions *is* visual perception.

Our explanation of the function of each particular cortical region differs somewhat from the detection of linking features described by Barlow in 1981 [16], but is supported by Barlow's more recent work. Because of the limited range of connections in the cortex, any particular cell only receives relatively local input. We suggest that like the coding system described by Field [71], neighbouring cells in any small region of cortex act to remove redundancy from their (relatively) common but local input. Because, initially the mapping from the many dimensional visual parameter space to the cortex preserves locality for the spatial parameters, it is only the redundancy defined in the visual data by these parameters that gets eliminated. The activity of the cells still exhibits redundancy for less local spatial relations and for all the other parameters in the visual data. But little by little, the redundancy in each of these is eliminated too, after non-topographic maps to further cortical regions.

7.12 Summary

The relationship-between-observations idea allows a glimpse of an aspect of the mechanism of perception, but not how these relationships come to exist. Several authors employ the notion of value or usefulness but it is difficult to see how this would be acquired. One of the ideas coming through in this chapter, is that as well as the fundamentally important level of description in terms of primitive observations, which is the main focus here, there seems to be a less abstract level of description – possibly an operational or a dynamic level of description on the scale of neural populations, where it might be more appropriate to investigate or model some aspects of perception. The observational level of description may be only an approximation to this operational level and may be too rigid to capture some of its features. The organisational closure description of Varela and the notions of circularities of description may be one suitable direction in which to continue this investigation.

Chapter 8

8 An Automated Visual Inspection Application

The topic of this chapter is not a direct development of, or application within, the theoretical framework which comprises the primary subject matter of this dissertation. It is rather, an attempt to show that the philosophical ideas discussed in chapter 2 are not completely divorced from real, down-to-earth problems. As such, it reinforces the message of chapter two, that there are alternative ways of addressing the issues of pattern recognition, objects and features, to the conventional, Aristotelian-influenced one.

8.1 Introduction

A problem with existing feature-based methods of automated visual inspection is the difficulty of selecting suitable feature sets. Usually the features used for error classification are heuristically determined. In this chapter we describe an attempt to improve difficult and repetitive inspection tasks by automatically selecting feature sets which best (in a well-defined sense) represent the distinctions required in a classification. The method is based on a modified version of the Karhunen-Loève transform (KLT) applied to sets of normalized imagelets of individual joints. The modification is a direct consequence of the "object-predicate inversion" ideas proposed by Watanabe as an alternative to more conventional pattern recognition approaches in certain cases. Watanabe's description of the theory involved is in terms of covariance matrices and this is reviewed below. We show how these ideas can be reinterpreted in terms of the Singular Value Decomposition (SVD) and illustrate the method by describing its application to the inspection of solder joints on surface-mount technology (SMT) printed circuit boards. The output of the coding method is a small coefficient set suitable for use with standard statistical classification techniques or with novel neural-network based classification. We describe how a simulation of a straight-forward neural-network classifier was used to classify the original solder joints with very good accuracy. This coding/classification implementation also gave us an opportunity to investigate the possibility of interpreting nodes within the network in symbolic terms.

The work described in this chapter is one aspect of a project which was originally motivated by the difficulty experienced in automating a particular industrial inspection process: the examination of solder joints on the legs of Surface Mount Technology (SMT) Integrated Circuit (IC) devices during the assembly of PCB-based products[1]. However, because of their generality the ideas described here have a much wider applicability. This particular application is simply one good example of their use.
Experience has shown that the best basis for a decision on the long term reliability of a solder joint is often *not* the electrical or mechanical properties at the time of manufacture, but the *visual appearance* of the joint[2]. It has however proved to be particularly difficult to describe the criteria that humans use to detect defects. Consequently efforts to reliably and robustly code these criteria in automated inspection systems using standard statistical image analysis has had only limited success. These facts suggest that an alternative approach might be of benefit.

This approach is described in the next section, 8.2. The first steps in this approach rely on the dimensionality reduction properties of the KLT, which is described in section 8.3.2. While the KLT is particularly well suited in theory to the type of application discussed here, there are major computational disadvantages. These disadvantages can be overcome by the implementation of a modified version of the KLT, which allows the eigenvectors to be found indirectly. This modified KLT was originally described in the Pattern Recognition literature by Watanabe [19] and more recently by Sirovich and Kirby [222,223]. The modified KLT and normal KLT are shown to be simply different facets of the SVD of the original data and this is described in section 8.3.3. The interpretation of the modified KLT in terms of the SVD allows for

---

[1]At the moment there is a growing sentiment within manufacturing industry that inspection of a product at the end of the production line does not add value to the product, in other words: "quality cannot be inspected into a product". Instead there is an effort to maintain much greater control over the production process and the production-line equipment itself. It is in this spirit that inspection is discussed here: not as an end point in the manufacturing process but as a sensing mechanism which is integrated at as many points as is necessary into the process, providing not just information about the product, but a feedback path for control of the process itself [5,7]. This type of implementation imposes a completely different set of requirements on the inspection sub-system(s) than finished-product inspection. In particular it means that many more inspection systems will need to be engineered, in shorter time, by people who are not necessarily vision engineers [6]. Not only does the automatic feature-set selection process described here allow problems to be solved which might not hither to have been possible, but allows this to be carried out in a very much problem-independent, and therefore not knowledge intensive way.

[2]Thermal and thermo-mechanical properties of joints are quite powerful as a basis for diagnostic procedures but are often considered unacceptably invasive by their nature. X-ray imaging is another direct way of determining the integrity of a solder joint but it can be quite expensive.

increased accuracy, particularly in limited word length applications. It also allows for the possibility of more efficient implementations based on modern algorithmic techniques for matrix decomposition though this part of the calculation is in any event carried out off-line. Finally, in section 8.4, examples are presented of the modified KLT/SVD being applied to a set of imagelets for the application concerned.

8.2 Pattern Recognition with Neural Networks

The term "artificial neural networks", (ANNs, or simply neural nets) is used to describe classes of non-linear algorithmic procedures and structures, loosely based on, or inspired by some aspect of natural neural systems. They are usually characterized by the massively parallel and distributed nature of their computational processes. In addition, neural nets typically provide a greater degree of robustness than traditional sequential Von Neumann-type computers, and an intrinsic potential to continuously adapt to the input data [224]. The most direct way of describing the function of a neural net is in terms of pattern classification. Here we take the broad view (see [19, p.199]) that the entire process of pattern recognition can be considered as a gradual, step-by-step reduction of the dimensionality of variables. The starting point is the raw observed quantities (pixels in our case) and the process culminates in a single variable with two values indicating membership or non-membership of a class (or generalisations thereof).

The structure of the problem concerned here is one of supervised adaptation to suitable decision classes. In other words, examples of both satisfactory and unsatisfactory versions of the solder joints are available for *off-line* "practice". During this, the system is expected to adapt to the criteria most suitable for distinguishing one set from the other. The *on-line* process is then required to use these criteria to automatically make the classification. There are indications that relatively simple neural nets, such as the multi-layer perceptron type, are capable of accomplishing this type of adaptation and classification – at least on data sets with quite small numbers of input parameters [224]. So while this type of pattern recognition is possible *in principle*, the large number of inputs (images of size $150 \times 50 = 7500$ pixels are used here) effectively rules out the direct application of neural nets with currently available technology.

8.2.1 Karhunen-Loève Preprocessing

In addition to the final classification step, the large numbers of inputs involved for pattern recognition on entire images, mean that some type of dimensionality reduction step must also take place. It is assumed that the input parameters (pixels in images) are not completely independent. They are correlated with each other and thus are dependent on some smaller number of unknown *underlying* parameters. The extent of this correlation is increased by taking care during the image formation and acquisition process to "pose" or register the individual images so that all unnecessary variation due to the position or size etc., of the item being imaged is removed [222]. Experiments with particular types of multi-layer neural nets applied to problems with a small number of inputs where there is some correlation between the inputs, have been carried out [18,225,226]. They have shown that the initial layers of the neural net adapt to effectively "code" the input data in terms of a much reduced number of parameters (dimensionality reduction). Different nodes within the networks often converge to semi-orthogonal combinations of the input parameters in a manner reminiscent of the KLT [18,225,227]. Indeed, Oja [228] shows theoretically that certain types of constrained adaptation in particular types of network lead the relevant network to estimate the eigenvectors of the input covariance, which is the KLT.[3]

It would be completely impractical to attempt to repeat these experimental results on data sets of the type under consideration here. The amount of computation for the large number of inputs, the number of images required for satisfactory adaptation, and the training time, are simply too large. These results do however indicate a way to proceed. If the dimensional reduction stage of the pattern recognition procedure were carried out by a standard algorithmic process such as the KLT, then the neural net might provide a powerful way of determining classifications on the basis of the output data from the KLT, with a much reduced number of parameters. A number of assumptions underpin this argument:

(i) that the preprocessing KL step reduces the number of parameters required to represent the input data to a number suitable for neural nets – of the order of tens rather than thousands or more,

---

[3]Note that this is not true of neural net in general.

(ii) that the remaining parameters are sufficient to *distinguish* between the required classifications, and

(iii) that a *linear* parameterization of the data is possible.

The first aim of the work described in this chapter is to use the linear analysis methods of the KLT to determine the number of parameters required to code the images to allow reliable and robust classification. The disadvantages described above of getting the neural net to carry out this step would thus be avoided, without losing the desirable discriminating properties of particular neural architectures.

8.3 Dimensionality Reduction

8.3.1 Features

As mentioned earlier, it is useful to consider pattern recognition as a process of step-by-step reduction in dimensionality. This reduction corresponds either explicitly or implicitly to making some variables more "important" than others. The more "important" variables are then referred to as "features" and can be used in the final steps of classification [229]. In general the only difference between "observations" and "features" is that there are fewer features, though they collectively contain most of the information required to make a classification on the basis of the original set of observations.[4]

A useful tool in the analysis of this type of pattern recognition problem is the mathematical concept of a vector space [230,231]. If each pixel is considered as an independent variable, then any image can be represented as a point in a space with as many dimensions as there are numbers of pixels. Thus a particular *150 x 50* image is a particular point in a *7500*-dimensional space. Most operations in this space are simply generalizations of the familiar vector and matrix operations in two and three dimensions.

[4]In Machine Vision, the term "feature" usually refers to some measurement taken on an image, such as distances, moments, areas, lines, etc. Jain [229] classifies features as spatial, transform, edge/boundary, shape, moments or texture on the basis of the type of processes used to derive them. Here we will be content with the more general idea of a feature described above. The term feature is simultaneously used on the one hand for single pixels, and on the other, for particular combinations of pixel values with the same dimensionality as the original images, ie. image-like combinations of pixels with particular properties

Consider, for example N images with M pixels each (ie. N objects, each subjected to the observation of M variables). We can express these as N feature vectors with M elements each: $\{X^{\alpha}_{i}\}$, where $i = 1,...,M$ labels components and $\alpha = 1,...,N$ labels

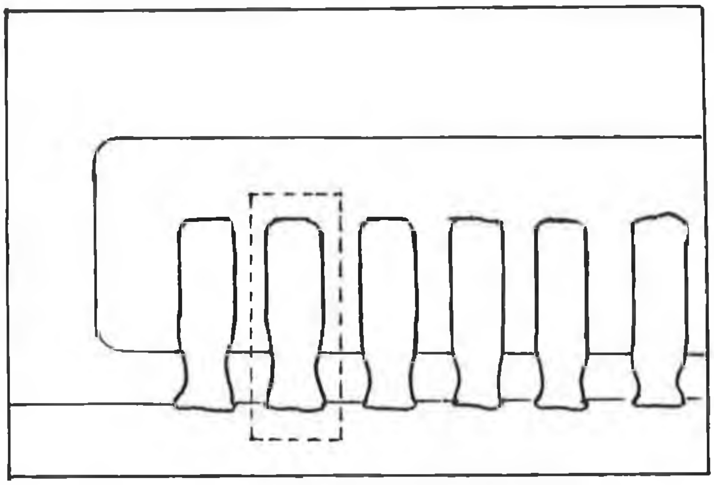


Figure 24. Schematic of SMT IC leg showing the positioning of a window used to "cut-out" the leg "imagelets".

vectors. In the case of general images, sizes are usually of the order of 512×512. Here we are dealing with segments "cut-out" from the original images (see Figure 24), referred to as "imagelets", which are typically of the order of 150×50, ie. $M = 7500$ pixels (see Figure 25). To make any solution practical, we need to reduce this to the order of $M' = 10$ to 100.

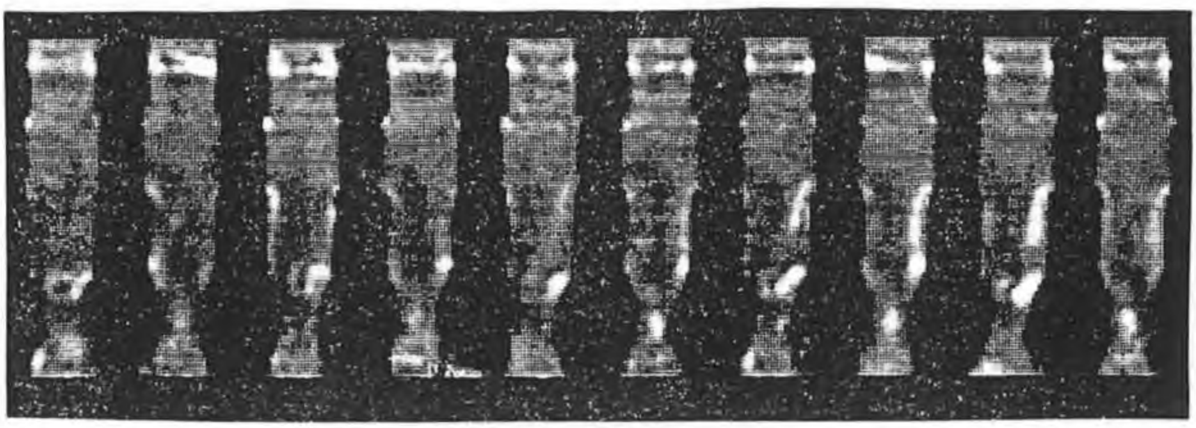


Figure 25. Ten examples of leg "imagelets" placed side by side for display purposes.

Consider, for example, N vectors consisting of two features (observations or components) each. In the arbitrary initial coordinate axes configuration shown in

Figure 26, the data are scattered in the plane defined by the orthogonal axes. In this contrived case the data seems to fall into two clusters, which might possibly be a useful basis for classification. Consequently the X_1 and X_2 axes are equally important or equally necessary in the classification process.

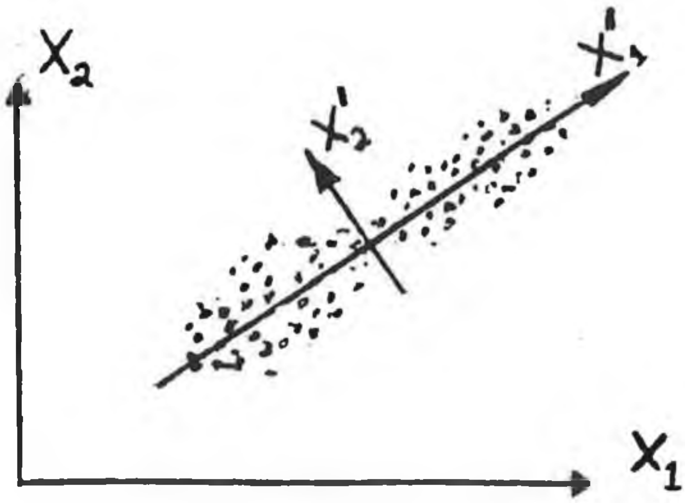


Figure 26. A schematic plot of a 2-D data set with substantial redundancy, but roughly equal variance in each variable.

If however we pick a different set of axes X'_1 and X'_2 as shown, then the X'_2 axis is now relatively unimportant for distinguishing in which cluster a particular point lies. That is, X'_2 is less important for representing the data, and by implication, as a basis for classification. In fact, components along the X'_1 axis alone are sufficient to allow a clear discrimination in this example. The dimensionality has been reduced from two to one and the classification step is correspondingly simpler. This is exactly the type of dimensionality reduction that we wish to achieve in our application. The only difference is that here our initial dimensionality is of the order of thousands or tens of thousands, while we can only cope with a number of variables of the order of ten to one hundred in the classification step.[5]

[5]This discussion somewhat over simplifies the situation by assuming that a good representation allows good classification. As Duda and Hart [230] and Therrein [231] state, this is not necessarily the case.

8.3.2 The Karhunen-Loève Transform

Formulation

Consider an $n \times m$ image or image sub-block. This is usually written and displayed as a 2-D matrix with n rows and m columns but for the purposes of mathematical analysis is often written as an $nm \times 1$ column vector (labelled here by X). The actual ordering of the vector is not important as long as the same ordering is maintained throughout. The KLT is based on the statistical properties of an ensemble of images of this type $\underline{X}^{\alpha}$. In general, α ranges over the elements of the ensemble, though if only a finite number of samples of the ensemble are available α is taken to range over these instead.

The pixel representation of an image is the standard interpretation of image data, where each pixel or variable represents the light intensity imaged at a point in the image plane from the original scene. In general, in the pixel representation or in an arbitrary unitary transform of it, the coefficients (pixel values or transform coefficients respectively) are correlated. A more efficient representation of the data could be obtained with a transformation which left the coefficients uncorrelated. This can be achieved by transforming to a coordinate system where the covariance matrix is diagonalized. The coordinate axes in this system are in fact the eigenvectors of the covariance matrix ([229] p.23, [60] p.127). The covariance matrix of the vectors $\underline{X}^{\alpha}$ is defined as $C_X = E\{(\underline{X}-\underline{\eta}_X)(\underline{X}-\underline{\eta}_X)^T\}$ where $\underline{\eta}_X = E\{\underline{X}\}$ is the mean vector and ^{T} denotes the vector or matrix transpose. For a given set of N samples from the ensemble, the mean vector and covariance matrix are approximated by

$$\eta_X = \frac{1}{N}\sum_{\alpha=1}^{N} X^{\alpha} \qquad C_X = \frac{1}{N}\sum_{\alpha=1}^{N} (X^{\alpha}-\eta_X)(X^{\alpha}-\eta_X)^T = \frac{1}{N}\sum_{\alpha=1}^{N} X^{\alpha}X^{\alpha T} - \eta_X\eta_X^T$$

Let e_i and λ_i, $i = 1,\dots,nm$ be the eigenvectors and corresponding eigenvalues respectively of the $nm \times nm$ matrix C_X (with the λ_i's in decreasing order: $\lambda_1 \geq,\dots,\geq \lambda_{nm}$). If A is the $nm \times nm$ unitary matrix with the vectors e_i as its columns, then A can be used as a transform to a coordinate system where the eigenvectors are the basis vectors. It can easily be shown that the covariance matrix $C_Y = A\, C_X\, A^T = \Lambda$ of the random vector $\underline{Y}$ given by $\underline{Y}^{\alpha} = A(\underline{X}^{\alpha}-\underline{\eta}_X)$, is diagonalized. Here Λ is an $nm \times nm$ matrix with the λ_i's along its diagonal and zeros elsewhere (see e.g. Gonzalez and

Wintz [139]). This transformation of the data vectors is reversible and the original data vectors $\underline{X}^{\alpha}$ can be recovered without error by means of the inverse transformation, i.e. $\underline{X}^{\alpha} = A^{-1}\underline{Y}^{\alpha} + \underline{\eta}_X$. Also C_X real symmetric $\Rightarrow A^{-1} = A^{T}$. The transformation by itself does not give any reduction in dimensionality because the transformed data vectors are still of dimension *nm x 1*, but the coefficients of the $\underline{Y}^{\alpha}$ vectors are uncorrelated.

Suppose that, instead of using all the eigenvectors of C_X, only the k eigenvectors corresponding to the k largest eigenvalues are used to form the columns of A (which will now be of size *nm x k*). The $\underline{Y}$ vectors will then be k-dimensional and the reconstruction given above will not be exact. The *mean square error* is

$$mse = \sum_{i=k+1}^{nm} \lambda_i$$

so the representational error is minimized with respect to all other possible orthogonal transforms coding to k coefficients by selecting the eigenvectors associated with the k largest eigenvalues. The KLT is optimal in a least-square-error sense over the sample set of data vectors selected by packing the most signal energy into the first k coefficients.

Computational Complexity

In the case of the *n x m* imagelets described above, the direct application of the KLT requires the construction and diagonalization of an *nm x nm* matrix. For $n = 150$ and $m = 50$, the covariance matrix has 7500^2 elements. For image processing applications of the KLT the *n x m* image array is usually divided into a number of equally sized blocks of size *p x q*. Each *p x q* block is coded as a unit (possibly represented as a *pq x 1* vector) independently of all other blocks. Also, the statistics are often assumed ergodic. The size of these blocks is usually in the range *8x8* to *16x16* pixels. This latter point is because correlations which exist between pixels and which the KLT is designed to remove, usually only exist over short distances in the neighbourhood of each pixel. Bigger blocks in general give diminishing returns due to the lack of a substantial amount of long range correlation. There are some notable exceptions to this point in particular applications and it is one of these exceptions that is exploited in this application.

Consider the case where all of the images in the ensemble are of the same *type* of object, with the objects all in register in precisely the same position in the image. The grey levels are also normalized to have identical grey level mean and variance. Then it is likely that whatever variation still exists between the individual objects that comprise the chosen set may be described by a small number of parameters. This is the case if say, all the images in the ensemble consist of human faces in register [222], or if in the case under consideration here, all the images are of a particular type of SMT IC leg and its solder joint to a PCB. Examples of some of these imagelets of IC legs are displayed in Figure 25.

In theory the ordinary KLT is capable of optimally reducing the original set of measurements or pixels to any required smaller number of parameters. It also directly gives the residual error introduced by representing with only this number of parameters. Unfortunately, the direct application of the KLT requires the calculation of the covariance matrix of the data vectors and its diagonalization. In the case of the IC leg imagelets (which are relatively small images of size $150 \times 50 = 7500$ pixels), the covariance matrix has dimensions of 7500×7500. Fortunately there is a way of carrying out the required dimensionality reduction without incurring impossible computational penalties. The ideas involved are most easily explained in terms of the singular value decomposition (SVD).

8.3.3 The Singular Value Decomposition

The Covariance Matrix by Matrix Products

Invariably, the covariance matrix is *defined* in terms of the outer products of $M \times 1$ vectors, and the linear sum of these outer products [139,229]. (Here $M = nm$ = number of pixels in the imagelets). However, the calculation of an approximation to the covariance matrix over a finite set of N samples from the ensemble actually amounts to the matrix multiplication of the $M \times N$ data "matrix" X^{α}_{i} with the transpose of itself, i.e. $(X^{\alpha}_{i})(X^{\alpha}_{i})^{T}$, or simply $\mathbf{XX}^{T}$ (see e.g. Golub and Van Loan, [232, p.10])[6]. We can use this fact as a bridge between the KLT and the SVD. Note that in our application each column of $\{X^{\alpha}_{i}\}$ is the vector form of a single image, i.e. each column has dimensions $M = 7500 \times 1$. If a sample of $N = 100$ images is used to calculate the

[6]We sometimes write the matrix $\{X^{\alpha}_{i}\}$ simply as X but it will usually be possible from the context to distinguish it from the random vector $\underline{X}$.

approximate eigenvectors, the matrix $\{X^{\alpha}_{i}\}$ will be of size 7500×100. (Note $\alpha = 1,\dots,N$ labels the columns of $\{X^{\alpha}_{i}\}$, while $i = 1,\dots,M$ where $M = nm$, labels the elements in each column, i.e. each row).

The SVD Formulation

The first step of the KLT is to get the eigenvectors of the covariance matrix XX^T. This operation is however reminiscent of the definition of the singular value decomposition (See e.g. [60, p.126; 229, p.176; 232, p.70]):

Definition: An arbitrary $M \times N$ matrix X of rank r can be decomposed into the sum of a weighted set of rank one matrices by the SVD. That is, there exist orthogonal matrices $U = [\underline{u}_1,\dots,\underline{u}_M] \in \mathbf{R}^{M\times M}$ and $V = [\underline{v}^1,\dots,\underline{v}^N] \in \mathbf{R}^{N\times N}$ such that $U^T X V = \Lambda^{1/2}$, where (for $\lambda_1 \geq,\dots,\geq \lambda_r$)

$$\Lambda^{\frac{1}{2}} = \begin{array}{c} \\ \uparrow \\ r \\ \downarrow \\ \uparrow \\ M-r \\ \downarrow \end{array} \begin{array}{c} \overset{\longleftarrow\ r\ \longrightarrow\ \longleftarrow\ N-r\ \longrightarrow}{\left[\begin{array}{ccc|c} \lambda_1^{\frac{1}{2}} & & & \\ & \ddots & & 0 \\ & & \lambda_r^{\frac{1}{2}} & \\ \hline & 0 & & 0 \end{array}\right]} \end{array} \qquad X = U\Lambda^{\frac{1}{2}}V^T = \sum_{p=1}^{r} \lambda_p^{\frac{1}{2}} (u_p)(v^p)^T$$

□

The columns of the unitary matrix U are composed of the $M \times 1$ eigenvectors $\underline{u}_i$ of the symmetric square $M \times M$ matrix XX^T. The columns of V are the eigenvectors $\underline{v}^{\alpha}$ of the symmetric square $N \times N$ matrix X^TX. The λ_i are the identical non-zero eigenvalues of both the XX^T and X^TX matrices. Since $N \leq M$, there are at most $r \leq N$ non-zero eigenvalues. The $\lambda^{1/2}_i$ are called the singular values of X and the vectors $\underline{u}_i$ and $\underline{v}^{\alpha}$, the i^{th} left singular vector and the α^{th} right singular vectors respectively. The outer products $\underline{u}_p\underline{v}^{pT}$ form the set of unit rank $M \times N$ matrices mentioned above. The SVD is also called the spectral representation or the outer product expansion of X. The row transformation matrix V performs the diagonalization operation $VX^TXV = \Lambda$, while the column transformation matrix U performs the diagonalization operation $UXX^TU = \Lambda$. The SVD has many useful properties [229, p.177]. Probably the most useful one here is the fact that if the r $N\times 1$ right singular vectors $\underline{v}^{\alpha}$ are known, the r $M\times 1$ left singular vectors $\underline{u}_i$ can be determined:

$$\underset{(Mx1)}{u_i} = \frac{1}{\sqrt{\lambda_i}} \underset{(MxN)}{[X]} \underset{(Nx1)}{v^n}$$

This is the crucial point of both the modified KLT and the SVD as it is applied here. Calculating the large *(7500 x 1)* left singular vectors is often virtually impossible in reasonably realistic problems of this type. But in fact they do not need to be calculated directly. They are available upon solution of potentially smaller problem.

Relationship between the SVD and KLT

One important aspect of the usual use of the SVD needs to be clarified at this point. Usually the SVD is defined for a *single* 2-d image, (or 2-d matrix in the general case [229, p.178]). The energy concentrated in the transform coefficients λ_p, $p = 1,...,k$ is maximized over *any* other unitary transform of that image or matrix. The KLT on the other hand maximizes the *average* energy in a given number k of transform coefficients where the average is taken over the sample images from the ensemble. On an image to image basis, the SVD concentrates *more energy* in the same number of coefficients than the KLT but has to be calculated for *each* image. The KLT needs to be calculated only once for the sample set and if this is representative of the whole ensemble, can even be used for any further samples of the ensemble.

The parallel drawn for our application between the KLT and the SVD is quite different from the general case described in the previous paragraph. Here the matrix decomposed by the SVD is a data matrix. Each of its columns is a vector representing a particular object. That is, *each* column vector in the data matrix could itself be an image. The entire set of N samples of images is represented in the single matrix X undergoing the SVD.

The SVD is not interesting here so much for its decomposition properties of X, but rather that it provides an alternative method for accessing the eigenvalues and eigenvectors of the covariance matrix XX^T. This can be done either via a direct SVD algorithm such as the LINPACK QR factorization algorithm used in Matlab™ [233,234,232] or using the eigenvectors of the other "pseudo-covariance" matrix X^TX. The latter case is particularly interesting for image processing and image coding where typically M is much larger than N. For example, *100* imagelets, each of size *150*

x *50* means that $M = 7500$ while $N = 100$. Solving for the right singular vectors by diagonalizing the covariance matrix means diagonalizing a *7500 x 7500* matrix. On the other hand, solving for the left singular vectors involves either implementing some type of reduced SVD algorithm (such as the "economy" SVD from Matlab) on the data matrix $\boldsymbol{X}$ or simply diagonalizing the *100 x 100* matrix $\boldsymbol{X}^T\boldsymbol{X}$. The eigenvectors of this matrix $\boldsymbol{X}^T\boldsymbol{X}$, which are also called the left singular vectors of the matrix $\boldsymbol{X}$ are *100 x 1* vectors. They can be used to directly calculate the originally required *7500 x 1* eigenvectors of the covariance matrix (the left singular vectors or eigenimages) of $\boldsymbol{X}$ by straightforward dot product operations using $\boldsymbol{X}$, as described above. Remember the *7500 x 7500* matrix $\boldsymbol{X}\boldsymbol{X}^T$ and the *100 x 100* matrix $\boldsymbol{X}^T\boldsymbol{X}$ have the same number of non-zero eigenvectors and have identical eigenvalues. Knowledge of one set of eigenvectors is equivalent to knowledge of the other, due to the ease with which one set can be calculated from the other.

8.4 Coding Procedure and Results[7]

The sides of SMT ICs, including the full length of the legs and the solder joints to the PCB were imaged live on a monitor as *512 x 512 x 8* bit images [235]. Each image contained at least ten legs. A graphics overlay consisting of *10* "windows" of size *150 x 50* pixels was superimposed on the live image. The PCB was placed so that the IC legs "fitted" into the windows, as shown in Figure 1, and the corresponding *7500* pixels for each leg was stored as an imagelet. A set of *100* images was then

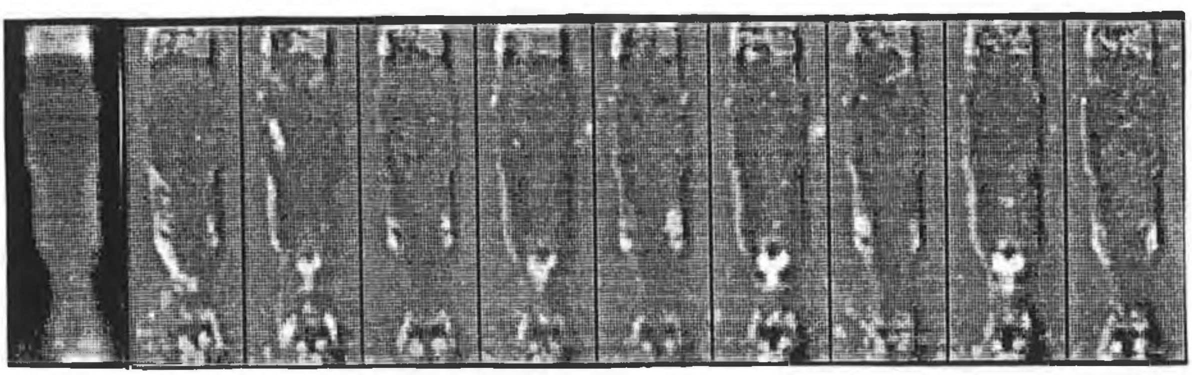


Figure 27. Mean image (first imagelet on left), and examples of "caricature" imagelets (original minus the mean).

picked out for processing. These *100* images were converted to vectors and loaded into

[7] The implementation of the coding and classification procedures described here was carried out by James Gunning as a part of his M.Eng. project under the author's supervision.

386-Matlab as a *7500 x 100* matrix. The mean column (imagelet) was calculated and subtracted from each of the *100* images to produce "caricatures" of the imagelets, as shown in Figure 27. The *100 x 100* matrix $X'X$ was calculated, diagonalized and its eigenvectors/values found. These *100 x 1* eigenvectors (right singular vectors) were in turn used to calculate the corresponding *7500 x 1* left singular vectors (or eigenimages). The first 10 eigenimages with largest eigenvalues are shown in Figure 27. Some test images were coded with the first 30 and the first 5 eigenvectors to yield 30 and 5 coefficients respectively. The test images were then reconstructed by linear combinations of the eigenimages. As shown in Figures 28 more eigenvalues give better results, but the main point is that images with *7500* pixels (variables) can be accurately coded with as few as 30 parameters.

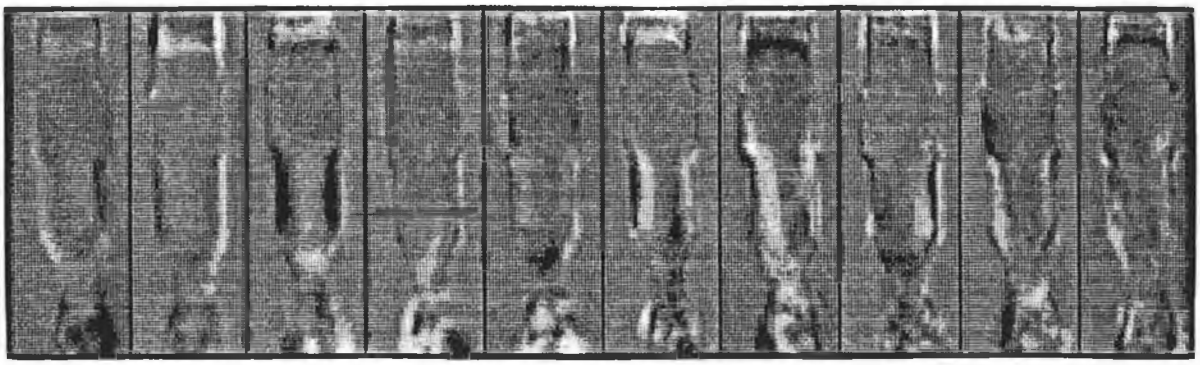


Figure 28. The ten eigenimagelets with the largest eigenvalues (decreasing from left to right).

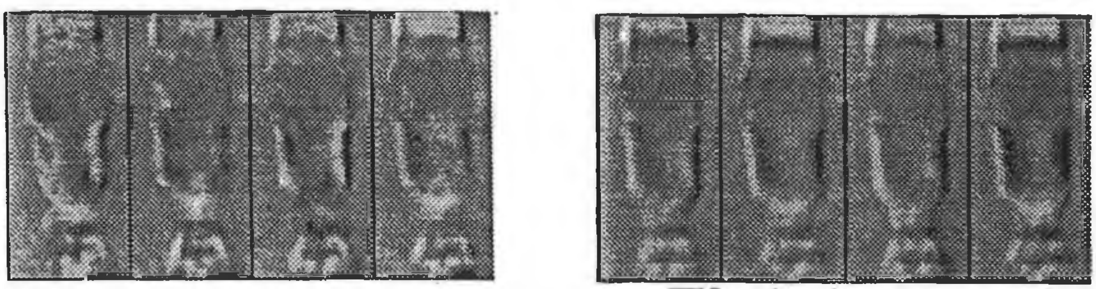


Figure 29. Caricature imagelets reconstructed with 30 coefficients (left) and with 5 coefficients (right).

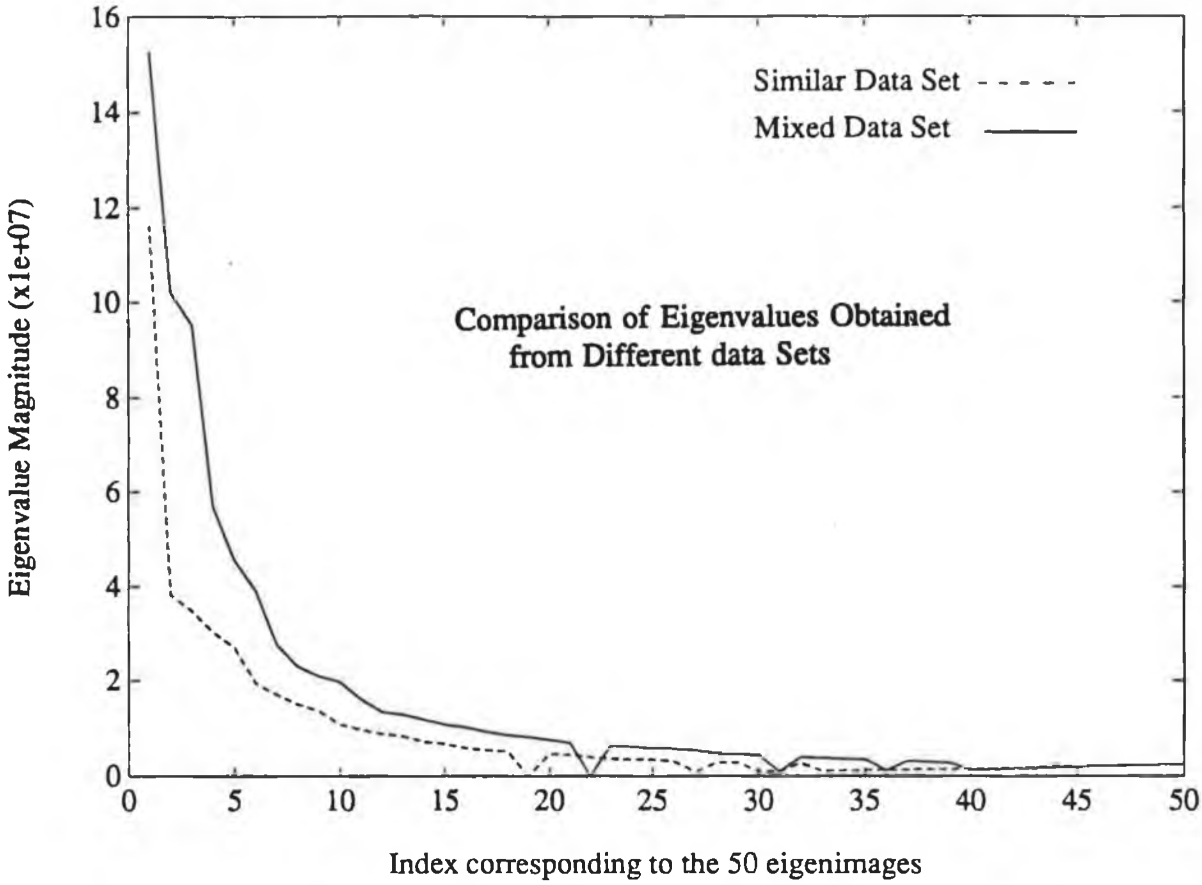


Figure 30. Plot of normalized eigenvalues vs eigenvalue index. Adapted from [235].

The *7500 x 1* and *100 x 1* left and right singular values were also calculated using the "economy" SVD function provided in Matlab. This took much longer, possibly due to the virtual memory swapping arrangement within 386-Matlab. With the IEEE floating point precision offered by Matlab, the results were virtually identical to those of the modified KLT. The eigenvalue data is plotted in Figure 30 and a graph showing the profile of two lines from original and reconstructed images is plotted in Figure 31. These results are only qualitative, and given purely to illustrate the ideas involved. Remember that the only part of this process that needs to be carried out on-line is the capture of test images and their coding with preselected eigenvectors. If for example 30 coefficients are required, then the computations involve 30 dot products of pairs of *7500 x 1* vectors.

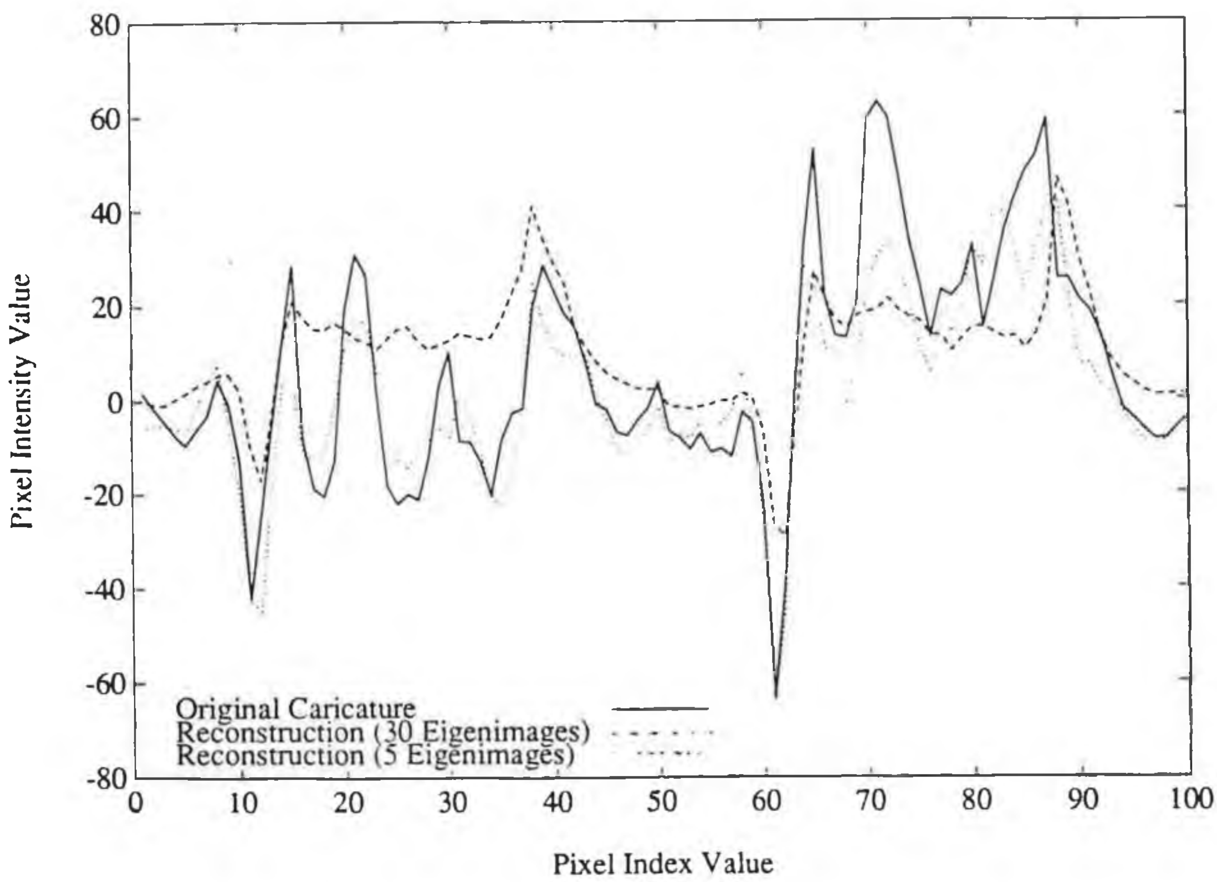


Figure 31. Plot of two succeeding lines (50 pixels per line) from the reconstructed caricatures. From [235].

8.5 Classification with Feature Sets

The third stage of the defect recognition process is to use the features or coefficients coded in the previous stage for classification. A multi-layer perceptron network structure with error-back propagation is used as the basis of the classification algorithm. This is augmented with a number of improvements to increase the rate and reliability of classification. Two of these, neural bias and adaptive training rate are reasonably standard and described in [236,237]. Some of the other improvements are less straight-forward and are briefly discussed below.

Momentum

This involves adding a term to the weight adjustment that is proportional to the amount of the "previous weight change". The weight adjustment equation is modified to become

$$\Delta W_{pq,k}(n+1)=\eta(\delta_{q,k}\times OUT_{pj})+\alpha(\Delta W_{pq,k}(n)) \quad (3)$$

using the notation described in [236]. α the momentum coefficient, is normally of the order of 0.9. Note that the "previous weight change" $\Delta w_{pq,k}(n)$ is the cumulative average weight change taken over the entire training set. This additional term increases the memory requirement by a factor of three but convergence rate is greatly improved. (Note: Convergence is quantified here using the root-mean-square (RMS) measure described by Dayhoff [236]. An RMS value below 0.1 is taken to indicate that the network has learnt the training set). As well as increasing the convergence rate the momentum term helps to avoid getting stuck in local minima.

Use of Noise in Network Training

The generalization properties of a neural net can be improved by adding pseudo-random noise to the training set. It also helps overcome problems associated with a small training set.

Pruning the Network

One major problem with ANN algorithms is choosing a network of the right size. The training algorithms tend to spread non-vanishing weights over the entire network, regardless of the size optimally required. There are a number of advantages to using a smaller network:

1. The time required to train and test a neural network grows linearly with the number of connections. Hence a smaller network is more efficient.
2. A neural network which is too large will simply memorise the training set and have poor generalisation ability. But if the neural network is too small it may never solve the problem.

Exhaustive testing of different sizes of networks is computationally unfeasible even for relatively small networks, so one solution is to prune a larger network [238].

Pruning involves estimating the sensitivity of the error function to the exclusion of each interconnection in the network. This is achieved by introducing "shadow" arrays that keep track of the incremental changes to the weights during backpropagation learning. The network is pruned by discarding the connections with the lowest values of sensitivity. The computation is relatively straightforward and the sensitivities can be estimated using the equation,

$$S_{ij}=-\sum_{0}^{N-1}\frac{\delta E}{\delta w_{ij}}(n)\ \Delta w_{ij}(n)\left(\frac{w^{f_{ij}}}{w^{f_{ij}}-w^{i_{ij}}}\right) \qquad (4)$$

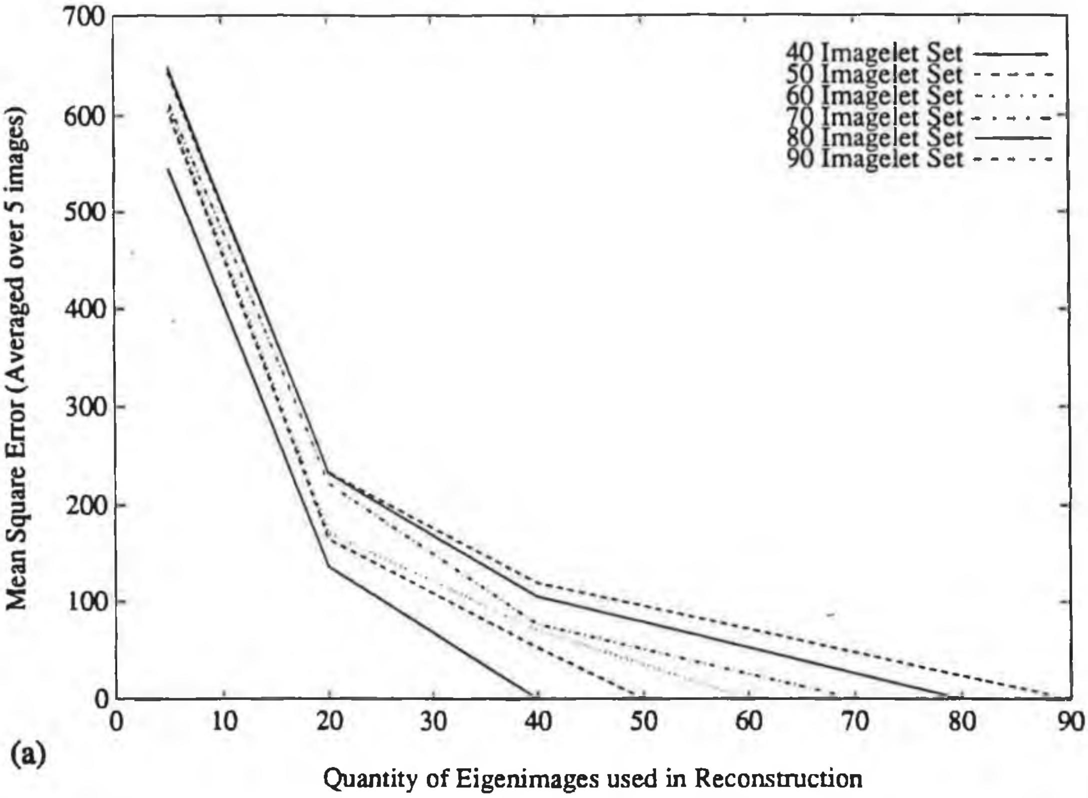
where S_{ij} is the estimated sensitivity of the error function to the removal of the weight w_{ij}, N is the number of training epochs, w^{i}_{ij} is the initial value of the weight and w^{f}_{ij} is the final trained value.

While there are other techniques for optimising network size [238], pruning provides a measure of the importance of inputs to the nodes in the various layers. This allows us to examine the relative importance of the coefficients from the coding stage for the classification process. It also provides the possibility of interpreting the functionality of the trained network in a rule-based sense.

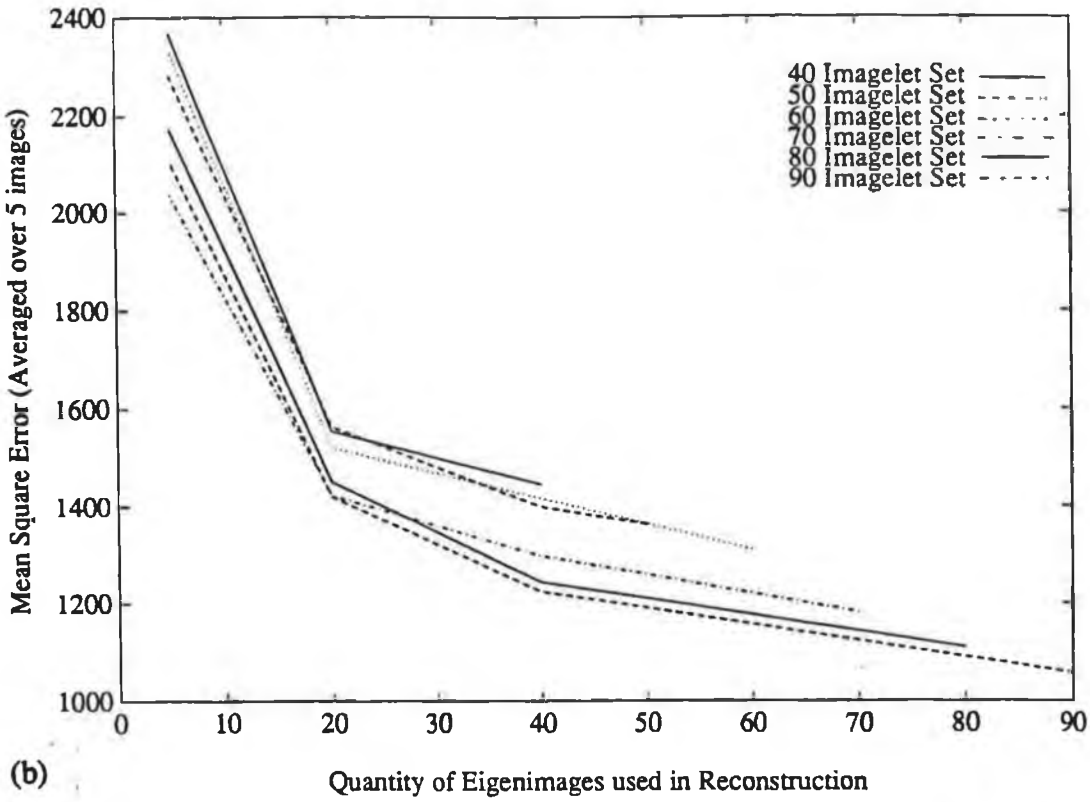
Coding and Feature Set Size

Figure 27 shows 10 eigen-imagelets corresponding to the largest eigenvalues and hence the largest variance. For imagelets external to the set used to construct the covariance matrix the KLT does not provide the best LSE representation. How representative the eigen-imagelets actually are of the global data set improves with the quantity and uniqueness of the imagelets used in their calculation, but with diminishing returns. Figure 29 shows this fall off clearly. Here we see that the difference between the average reconstruction error using 20 eigen-imagelets from a *90x90* covariance matrix and a *40x40* matrix is almost negligible in this case.

On the basis of these results 20 eigen-imagelets with the largest variance were used to represent the imagelets initially. In other words the neural network classifies imagelets on the basis of 20 coefficients only. For initial testing the eigen-imagelets were obtained from a *50x50* covariance matrix. To ensure the eigen-imagelets were representative of a global data set, the 50 imagelets included in the covariance matrix calculations were an admixture of good solder joints, and defective joints from the classes with excessive solder, insufficient solder, bridging and displaced legs. The coefficient sets obtained in the coding stage were then used in training and testing the neural network.



(a)



(b)

Figure 32. Plot of average reconstruction errors for "internal" images (a) and for "external" images (b).

The Neural Network Configuration

The number of input nodes is usually decided by the data preprocessing stage, one node for each input variable. The number of output nodes is task dependant, determined by the number of classes - five in our case. The problem is to decide on the number of hidden nodes or layers.

Three active layers were chosen to allow the most general type of classification, if required. The issue of the number of hidden nodes is discussed in the literature [224,239,240] but no clear rules are available for a particular case. After some experimentation the numbers were chosen as 12 nodes in the first hidden layer and 8 nodes in the second hidden layer, ie., a 20:12:8:5 neural network. It does not matter if this is in excess of the optimum requirements, because of the subsequent pruning.

Simulation of the Neural Network

The coding stage and ANN were simulated with the matrix processing package Matlab™. Random noise with a uniform distribution was added to each coefficient (in proportion to the coefficient size) to aid the convergence and generalisation of the network. The neural network typically converges to an RMS value of below 0.1 using an adaptive training rate in approx. 50 iterations. Each iteration consists of a single pass through the network with a 50 imagelet training set. The trained network was then tested using coded imagelets that were external to the training data and the result was a 100% correct classification rate.

Optimising the Neural Network Size

The sensitivity results obtained from the 20:12:8:5 network showed that the majority of the large sensitivity (important) connection weights from the input layer to the first hidden layer were from the coefficients corresponding to the eigen-imagelets with larger variance. However, none of the nodes in the first hidden layer had important connections from the input coefficient with the largest variance. A plot of the coefficients corresponding to the eigen-imagelet with the largest variance is shown in Figure 33. The traces show examples of the range of coefficients' values corresponding to the different classes of solder joints. They show that this coefficient is important for distinguishing displaced solder joints from the other four joint types but does not distinguish between these other types. It is because it does not distinguish

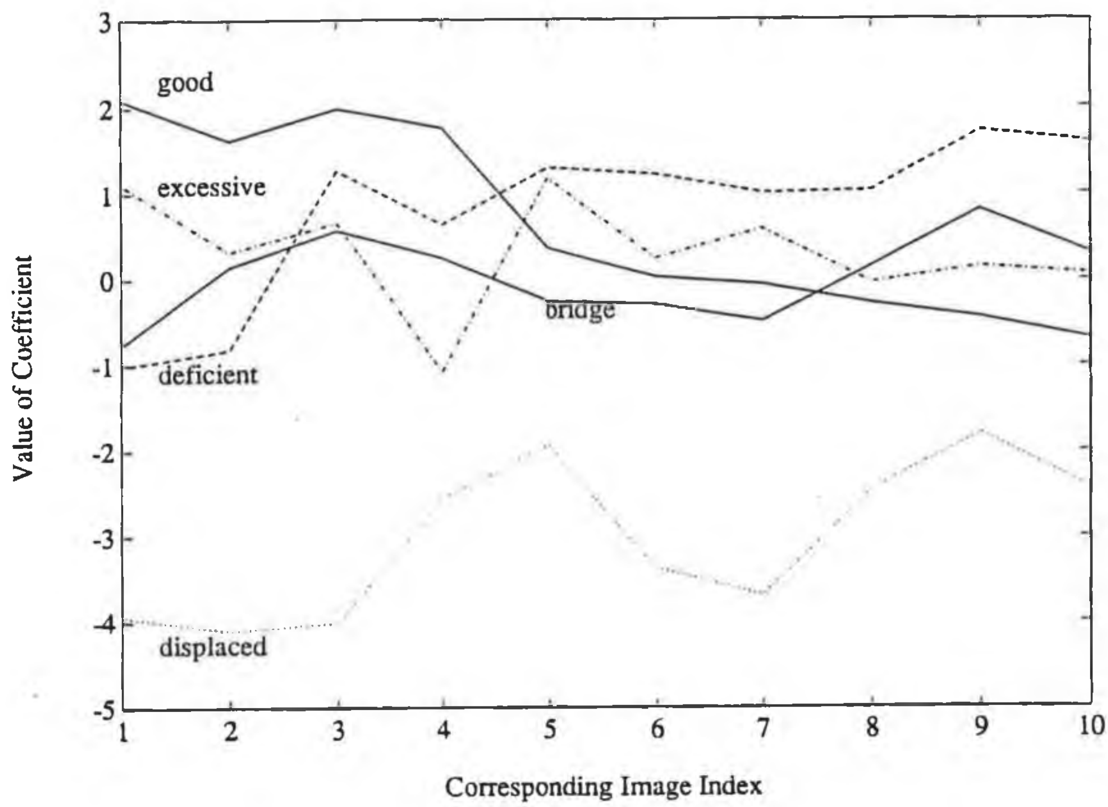


Figure 33. Plot of the coefficients corresponding to the eigenimagelet with the largest variance. From [235].

between these four types that it consistently shows low sensitivity.

The overall results of this pruning process was to reduce the neural network from a 20:12:8:5 to a 7:7:7:5 configuration with only a slight decrease in classification robustness. This meant that classification of image content with as few as 7 coefficients (down from 7500) was possible. The pruning implemented was a modified version of the original algorithm which was designed for pruning individual low sensitivity weights. Our main concern is with nodes - particularly input nodes - so entire nodes with low sensitivity weighted inputs were pruned out and discarded.

Some indication of the features the nodes in the neural network extract is shown in figure 34. Each image corresponds to a single node, i.e., 7 nodes in the first hidden layer, 6 in the second and five output nodes. Each image is calculated by averaging all images that activate a particular node and subtracting the average of the images that inhibit the node. The result is a picture of the internal representation of the ANN.

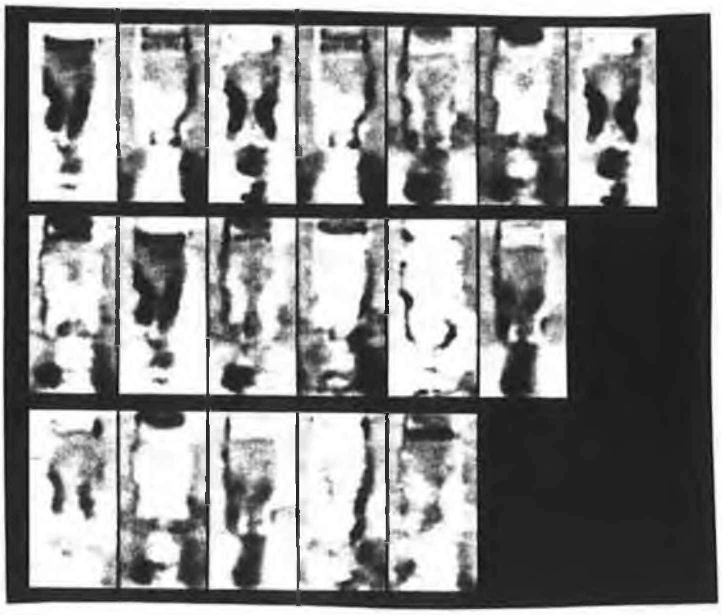


Figure 34. Characteristic images of each of the nodes (7:6:5) in a three-layer network. From [235].

8.6 Application to Motion Recognition

The ideas described above of KLT-based dimensionality reduction or coding and neural-network-based recognition has also been applied to the problem of recognising people through their particular posture or gait during normal walking motion. The experimental setup used was particularly simplified in order just to demonstrate the feasibility of coding/recognition idea. Three sequences of different people walking on a treadmill were used (30 seconds each at 12½ frames per second). Each image was reduced to 32x64 pixels, to reduce computational overhead, but also so that the people would be less likely to be recognised by some detail not relevant to the motion. At walking speed, the people went through one complete walking (pose) cycle in 14-15 seconds. Four pose cycles were taken as a representative set for the coding and recognition stages (56-60 images). The covariance matrix was constructed as above without taking cognisance of the order of the images in the original sequences. It was found that any of the images could be represented with a 5% error for 20 eigenimages constructed from the same sequence and to within 8-12% error for images not from the sequence used to construct the eigenimages.

Two methods of analysis were used on the coefficient data. Firstly the Fourier transform was taken of the coefficients for each image in a sequence. This clearly showed the periodicities associated with the movement of the legs, arms and body during walking. The second type of analysis was to use a neural network to learn the particular position in a pose cycle corresponding to the coefficients of the image of that position in the cycle. The output of the neural net (a pose cycle position value) was then used to control the pose of a graphical "stick figure". The pose positions of the stick figure were derived by associating a particular angle for each of the major body joints with a position in a walking pose cycle. The computer graphics sequence generated using the cycle position output by the neural network was then animated into a video display. The result was that the stick figure quite accurately followed the motion of the person in the corresponding images. No attempt has been made to implement a person recognition as opposed to a pose recognition process as yet, though the results from the pose recognition were very encouraging.

8.7 Summary and Conclusions

The results were very encouraging and worked well for both of the applications described. They indicate that for given situations KLT preprocessing can be used to substantially reduce the amount of data required for classification of image content. The online operations required to carry out this preprocessing are simply the multiply-accumulate operations of the dot product and are easily implemented on existing hardware. ANN's can then be used to provide a robust classifier that also allows further selection of the important inputs. Work is ongoing to investigate the possibility of integrating this distributed decision approach with more traditional rule-based heuristic approaches in the industrial inspection application and more general pose and person recognition processes in the motion analysis application.

Chapter 9

9 Conclusions

Originally, at the start of the project which has led to this dissertation, we saw vision as a very useful capacity with which to equip a robot manipulator, particularly when we expected the manipulator to operate in a relatively unconstrained environment. We did not set out to prove any particular thesis about how this *should* be done, but sought to find a way that it *could* be done. In other words we wanted to find and use some suitable "theory of vision", or some appropriate formalism, that would allow us to understand the processes and problems of vision, that would allow us to design a vision system with particular goals in mind. After extensive research through a broad range of literature, after extensive efforts to understand and describe visual perception, to come to terms with what really are the problems with vision, and to see the scope for potential solutions to these problems, we have come to a position which partially answers these questions: at this stage it appears that there are at least three different ways that we can approach vision problems. These are the conventional representational approach, the radically different enactive approach, and the primary topic of this dissertation which is the an information theoretic approach.

The Representational Approach

One possible approach to understanding and working with vision, is the standard representational approach. The basic assumption here is that the environment of our machine is sufficiently well defined or constrained, that we can capture it in a suitable type of representation, and that we can give our machine enough information to manipulate and update this representation as it requires. We firstly (possibly explicitly but more usually implicitly) determine a symbolic representation of *our* world. Then, based on the assumption that this representation or description is also an appropriate description of what we see as the system's environment, we "ground" the symbols in the system's formal processes, by providing the appropriate links or heuristics. In this context, information should be interpreted as fixed and instructive, and we are ultimately the creators of this information. It is possible in the constrained circumstances characteristic of this approach, for us to formulate (at least in principle), an explicit symbolic representation of the environment of the machine in question, with a level of approximation which is constantly bounded[1]. Because of this, there is a direct sense in which the amount of information required to describe the situation is fixed. Furthermore, our use of this information is instructive, in the sense that the machine which uses this information is, in a very literal way, just carrying out our instructions.

The actual means of instructing the machine within this approach, may be made increasingly flexible by the use of sophisticated interfaces, such as natural language interfaces or voice activation, or alternatively by using sophisticated programming languages like prolog and programming techniques like sub-symbolic computation as in artificial neural networks. Nevertheless, we must be careful to always acknowledge the special role of the designer, programmer or engineer as the ultimate source of instruction – the grounding for the system's information or symbols. This role is thus one of an *observer*, separate from the system, whose function is to provide the appropriate mechanisms, the appropriate computational processes, and the appropriate connections or definitions for the system's symbols. In this sense the major thrust of the recent upsurge in interest in connectionist approaches for information processing is not qualitatively different from the more established computational approach which is explicitly representational. The basic epistemological position of much of connectionism is still representational by design. Of course there is absolutely no problem with this, as long as the fact is realized. The majority of artificial neural systems are simply trained to do exactly what we teach them to do, and often do it quite successfully.

That this, fundamentally representational approach is immensely valuable, is illustrated by the continuing success of machine vision applications. Machine vision works, not because its representations are necessarily identical to what we see as our picture or representation of the world, but because the representations and processes used are geared to the problem, and the problem is relatively tightly constrained. This approach is valuable as long as we recognise the assumptions that are implicit in it and which determine its limitations. The representations and processes of machine vision are still representations and processes constructed by us, or derived from our heuristics. The

---

[1]See the discussion on Rosen's meaning of the terms *simple* and *complex* in one of the footnotes in section 2.2.3 above.

success of machine vision will always only be as good as our understanding of the problem or the implications of our heuristics[2].

This approach on the other hand, is *not* suitable for describing biological systems, and it is not suitable for designing truly autonomous systems capable of displaying anything like the level of "intelligence" or "common sense" that we expect of humans. In other words it is not suitable for many of the applications or situations usually associated with general computer vision systems, where the emphasis is on unconstrained environments, behaviours and interactions. Now, whatever the merits of the representational approach, there was implicit in our research from the very beginning, a desire to consider relatively flexible and unconstrained situations as the environment for the visual processes that we were investigating. Thus, in hindsight, we can be quite sure that representational approaches are not the way to ensure success and we are forced to consider alternatives.

The Enactive Approach

The second approach to understanding the problems posed by vision, and indeed by perception in general, is related to a subset of the connectionist strategy for adaptation and learning, and also has its roots in efforts to understand the essential nature of the meaning of "living". In fact it is a confluence of the work of researchers in areas ranging from philosophy, theoretical biology, robotics and connectionism, which has in common the rejection of the representational position as a means of understanding the operation and behaviour of real cognitive systems. The work of several researchers in this area has been discussed in the body of this dissertation, but pre-eminent for its absolute incisiveness and all-encompassing logic is the extensive work of Humberto Maturana and Francisco Varela. Their work is usually associated with the term *autopoiesis,* though this is only one particular case of a much wider view that they express on the nature of explanation, the relationship between observer, machine and environment, the autonomy of living systems, and so on.

---

[2]It is worth commenting that at the moment, neither of these factors are the primary constraints on what level of penetration it is possible to achieve with machine vision in industry. The effort and expertise required to engineer each individual application with current hardware and software technology are generally more immediate constraints [6].

It is impossible to do justice to this approach here by trying to explain what is a quite radical reappraisal of the nature of being and knowing, strongly at odds with the mainstream of objectivism. Nevertheless I try to make what are the major points that differ from the representational approach above. From the perspective of this approach, for which Varela has recently coined the term *enactive*, information is literally *in-formation* – something that is formed within, something that is constructed. The "meaning" of information, so to speak, is related only to the continued viability of the system's functioning. Information is never picked up or transferred, nor is there any difference between informational and non-informational entities in the system's environment. Any significance that can be attributed, from the perspective of the cognitive system, to aspects of *its* world, arises solely from within the system itself. This is not in the solipsistic sense, meaning that significance is attributed arbitrarily and independently of the physical substrate of the organism's existence, but in the sense of the significance and signifier arising in a process of mutual specification. Neither the structure of the cognitive system's world, nor the operations of the cognitive system as observer are pre-given – they are co-determined by a history of cognitive interaction, neither logically preceding the other, but still logically compatible. What is significant – what can be known – is *enacted* or brought forth from an un-differentiated background. The fundamental logic or "foundation of reference" of the enactive approach is neither reductionist nor holistic, but an intermediate position which emphasizes the intrinsic circularity or reciprocal causality of cognition. Instead of perception being considered as primarily for the control of action, which is the conventional view, in the enactive view, action or behaviour is considered mostly in terms of the control or regulation of perception.

The only things relevant to the cognitive system's ontology in the enactive perspective are its organisation, its realization of that organisation in terms of a particular structure and the maintenance of its organisation. The relationship between the observer, the system, the system's environment and the system's identity are all made clear to avoid confusion, particularly confusion about the source of signification (or meaning). There *is* a symbolic role for information in this context, where an observer (or observer community) can decide to use a symbol as an abbreviation for a chain of nomic links relative to the organisation of the system. But this assignment of a symbol makes no sense outside the context of the organisation of the system and it is also not operational for the system itself. This is the case, because the symbol is defined by the observer who is in the privileged position of being able to interact with *both* the systems and its environment, but who suffers from the disadvantage that they cannot see what they think is the system's environment from the system's point of view – what the system sees as its *world.*

An example from autopoiesis of the type of circularity typical of the enactive approach is the so-called genetic "code". The meaning or significance of, or specification by, the base sequences of the genetic "code", makes no sense outside the context of the metabolic machinery which interprets that code to regenerate itself[3]. Neither the genetic code nor the metabolic machinery are logically prior to the other. They are co-dependent and exist by mutual definition. Similarly in the cognitive domain we again have an example of the pervading circularity, not in a paradoxical sense, but in the nature of the definition of the phenomenon itself. The world that we perceive is not a particular objective world waiting for us to open our eyes and to look at it. What we perceive as *our world*, along with our ability to perceive, have arisen by mutual definition, in the history of each of our individual interactions with what is not us in the medium of our realization. We actively determine what is important, what we want to see, in a way that gives us an illusion of a stable solid world that could, it seems, not be any other way.

For a variety of reasons, the enactive approach, (or what has also been called autopoietic theory) has had little direct success in computer or information technology applications, including vision, to date [241]. The primary reason is that the ideas were never directly aimed towards addressing engineering issues: the principal aim was to understand the grounding or basis of biological systems and cognition. In fact the main reason for any change in this, is not that the emphasis in autopoietic theory has changed very much. It is that in research areas like computer vision and artificial

[3]Seeing that the genetic structure is stable through many generations of reproduction and ongoing metabolic regeneration we might refer to the chain of nomic links that involves this structure, and gives it stability by some label or symbol. But this is an arbitrary reference made by us from our viewpoint. It does not describe the dynamics of the system which provide the nomic reason for the stability of this particular structure. Note that we can still describe the components involved in the genetic structure in information theoretic terms, and describe their capacity to be a link in the specification of regularities over many generations, in terms of an information channel capacity. However, their meaning still comes from the entire organisation: genetic and metabolic components.

life, engineering has begun to address issues which are normally proper to the domain of biology and cognition.

The work of Brooks in an alternative approach to robotics, is however, an example in the spirit of the enactive approach, even though it is not formally within the framework described by Maturana and Varela. Brooks' approach is to eschew representations altogether, preferring to let the world "serve as its own model". The robotic designs are not based on serial processing in stages or blocks called say, "sensing", "planning", "action", which has been the traditional robot control paradigm. Rather the complete sensory/motor control system is designed for a primitive task or functionality like walking. Many of these task-systems are then combined together in what is referred to as a *subsumption* architecture. This work is presently at an early stage of research and development. While there has been some success with the particular systems designed, there are major engineering problems in designing systems for more complex behaviours. At the workshop on *Autopoiesis and Perception* held in DCU in August 1992, there was some pessimism expressed about the possibility of being able to engineer systems with "usefully" complex behaviours. It was believed that the process of generating "better" machines within the general *enactive* paradigm, might be more related to animal husbandry, than to the conventional process of engineering and design. Whichever of these two extremes the final situation is nearest to remains to be seen, but the first steps towards this would be to develop a suitable mathematical formalism for setting up models of the concepts involved and making predictions about what might be possible.

The autopoietic theory or enactive approach stresses the organisation of a system, in terms of an operational description, as the embodiment of its properties, or abilities, etc. Varela has proposed an algebraic system for describing the abstract recursivity or self-referentiality inherent in the circular definition of autopoietic and cognitive systems, but little work seems to have been carried out using this formalism in practical situations to date. A less general, but more accessible level of formulation is in terms of dynamics. However, the sensory signal processing properties of dynamical systems is a subject which is only beginning to be explored. So all in all, the enactive approach as a theoretical paradigm is impeccable, but much work needs yet to be done to realize concrete benefits in practical applications. The third approach also attempts to highlight problems with the conventional view of computer vision, but does so in a less radical way than the *enactive* approach. It is this that we turn to next.

The Information Theoretic Approach

A common thread through the first two approaches described above has been the different ways of viewing information. In the third approach, information takes centre stage as the primary subject and tool of investigation. This third approach is the main topic of this dissertation. As enunciated in the abstract, the central motivation for the research leading to this dissertation has been to determine what we should be doing in our research into artificial vision systems — exactly what problems should we be trying to solve. This dissertation is an attempt to show that we can at least begin to answer these questions in the context of a quantitative theory of information.

The notion of "closing the control loop", coupled with the desire to understand how to build machines which could learn and behave in unstructured environments set the opening agenda for our research. Although it was not initially apparent, it is now clear that the representational approach described above would not be suitable for this type of situation. In the early stages of the research, literature on biological vision ranging from neuroscience to cognitive psychology was used to try to give a better understanding of the issues involved. After a time it was realized from this literature that there were deep-rooted problems in understanding visual perception, either biological or artificial, in terms of the conventional representational approach. A part of this process was the realization that there is much more in common between biological and computer vision than there is between computer and machine vision. Machine vision was seen to occupy a world of fixed information, controlled variation and restricted representations — a man-made world where man-made ideas about the world are perfectly useful. But at a certain stage it became apparent that these ideas were no longer valid within the research agenda that had been adopted; that they cannot be the basis of explanations within biological vision, and that they cannot provide a suitable basis for examining the issues of general computer vision systems or autonomous systems.

In general, the processes and representations involved in machine vision are relatively accessible. The decisions to be made automatically by the system, and the analysis or heuristics leading to them, are usually formulated explicitly. One of the main issues that arises on the realization of a fundamental shift in perspective between machine and computer vision, is to find a suitable theoretical framework and mathematical formalism for describing the development and operation of biological vision systems in a way that would be common with the related computer vision capacities. On the other hand, it must also be realized that it was not good enough to simply copy biological processes and functions. If we were to be able to use results from biological vision in any meaningful way, we would need to have some fundamental understanding of what problems were being solved by visual sensing and sensory systems. A suitable theoretical framework for dealing with these issues is presented and argued for in this dissertation. The development of an appropriate mathematical formalism, based on the fundamental groundwork laid here, is the next step. We already have some clues as to how we should progress in the work of Rosen, Varela and Wilson.

There were three main topics of influence which helped to put some structure on the approach to solving the problem of finding a suitable theoretical framework within which to work. The first was the anatomy and physiology of the early visual system, particularly (i) the mappings (topographical and non-topographical) between different visual areas, (ii) the pathways through the visual system for separately processing form, colour and motion/depth[4], and also including (iii) the massive reafferent projections in the "opposite" direction to the expected direction of information flow. These mapping or projection ideas, and the multiple pathways ideas are largely not represented in this dissertation. This is because they are not central to the conclusion presented here that information theoretic concepts are the appropriate means for answering many of the questions posed about the nature of vision. They would also have made the dissertation unmanageably large, and biased it towards biological vision, which was not the intention. They have however informed a particular way of understanding the nature of vision which is a subtext for the primary presentation.

---

[4]The idea being alluded to here is the magno/parvo separation described by a number of research groups working in neuroscience, particularly the work of Livingstone and Hubel. On the basis of anatomical, physiological and psychophysical evidence they identified three pathways through the early visual system of primates which are at least *somewhat* specialized for the separate processing of shape and fine detail, colour, and 3-D organization including depth and motion. The basic dichotomy here is like the original "what" versus "where" one identified by Marr. More recent work by Goodale and Milner has suggested that a "what" versus "how" dichotomy might be a better characterization.

The second direct influence was the use of information ideas and information theory to describe and explain a wide variety of visual concepts and phenomena, ranging from the processing stages in the retina of the fly, to the uncertainty relationships between various parameters in visual analysis. This of course has become central to the research approach or theoretical framework which is presented here as an appropriate way of dealing with the issues arising in vision, either artificial or biological. Now there are at least four distinct interpretations of the term information, and attempting to distinguish and relate these different views has been a part of the research methodology used here to support the primary aim of understanding perception. This for instance, is the subject of chapter 4 where the quantitative concept of information is traced from its roots in physics and engineering to the position presented here of using it as the basis of explanations in biological sensing systems. Prior to this, in chapter 3, an extensive survey of aspects and examples of biological vision is presented, to illustrate the extent to which much of what is known about biological vision can in fact be explained in information theoretic terms alone.

The third major influence which led to the theoretical framework for understanding perception which is presented here, was an analysis of the pattern recognition aspects of perception and a philosophical reinterpretation of the roots of pattern recognition, which is described at some length in chapter 2. Having assumed that it makes sense to talk about perceptual systems, the next step was to try to discover or describe the processes or properties that allow a system to be described as perceptual. This careful reexamination of the nature of pattern recognition led to the conclusion that it is possible to describe a fundamental perceptual primitive – the basic unit or event or process which is essential in order to describe a system as perceptual. Furthermore, we make the claim that this fundamental level of description can be effectively applied to all sensory modalities in particular organisms like ourselves, and also to perceptual capacities in different organisms, though more evidence needs to be gathered to fully validate this conviction. This is a property that is essential to any attempt at a unified treatment of perception. The human nervous system for example, does not build 3-D representations for vision, or pitch/location representations for sound, because these are the particular solutions of computational problems in visual perception or aural perception respectively. We do not pretend to understand why our conscious perception distinguishes as it does, between different subjective aspects of perception within one sensory modality, or between different modalities. But we do claim that, in terms of the processes or mechanisms of perception as they are currently understood, there is nothing to distinguish different modalities, other than say, the intrinsic dimensionality or statistical properties of the sensory signals, or the type of transduction involved, or the gross genetically specified architecture of the corresponding neural systems. Acknowledging that there is a level of genetic programming in the gross architecture of the nervous system, which affects the way its properties are expressed, it is still possible to claim that the primary determinant of the development and operation of a perceptual system is the same across modalities, and can be explained in terms of information and measurement theory alone, without recourse to representational notions of modelling aspects of the world.

With the benefit of hindsight it is possible to see that progress in solving these problems was only possible in the context of the shift of emphasis away from representationalism and objectivism, and this shift was prompted by the effect of these three areas of influence on the development of my understanding of the problems. This new way (for the author) of looking at visual perception is based first and foremost on the realisation that how we see depends on us, and our perceptual system's capacities, and on its history of visual experience – particularly at certain critical times during development. It is based on the realisation that we cannot "see" objects: we construct (some might say hallucinate) "objects", on the basis of interpretations of light signals, based on our structure and our prior experience. Furthermore we claim, on the basis of the treatment of the philosophical foundations of pattern recognition described above, and on the basis of the understanding of the notion of *relative information* and *signal-to-symbol* transitions in the context of information theory, that the most primitive signal transformation which can be of any relevance to an organism interacting with its environment is just such a signal-to-symbol transition, in other words, a primitive measurement or classification process.

Overall this approach is based on the realisation, that in order to allow something we design to escape our particular conception of reality, we must not try to incorporate our interpretations of the functional capabilities of our visual system. There must be a freedom for the system to find its own relevance in its own environment with its own organisation. The highest level at which we can describe this, without encoding our observables, or our values, is, I submit, at an informational level: at the level of particular information-carrying signals relative to a source and relative to a receiver, and at the level of ensemble averages of information rates, equivocation and so on. This level of description is not intended to be incompatible with the *enactive* one. It is intended to be a level which captures some of the important properties of the underlying system's organisation, not to replace it, but if necessary to generalize it to other realizations and other mechanisms or structures. It is not intended either to be directly operational for the system, but one might find aspects of the system's dynamics, and particularly the system's organisation, which are reflected in properties of the informational description. Thus for example the attractor states described in dynamical models of cortical operation and the capability to quickly switch between different attractors seems to have a parallel in the primitive decisions or measurements of classifications postulated as one of the fundamental components of the informational description.

The main justification for the informational approach presented here, as opposed to the purely *operational* of the enactive viewpoint, is that we need to get "low" enough to find common explanations across many different realizations of the visual capacity. By explaining biological vision we *are* explaining vision, as long as our explanations are at a level where they are not trying to interpret what are actually accidents of evolution.

Both the results presented by Linsker and Marshall on the self-organization of perceptual networks, and the view that the fundamental unit of perception is a primitive observation or recognition event, can be discussed in terms of information theory. It is not yet clear however, if both sets of ideas are at precisely the same level of description. The more obvious interpretation is of signals converging to nodes, and the decision or classification operations taking place at these nodes in the manner of a generalized perceptron, say. A different, and the author feels more plausible view is illustrated in Figure 34. Here there is not a unidirectional flow of signals from "input" towards "output" in the manner of a multi-layer feed-forward perceptron network. Instead, the basic operational unit is a bidirectional loop or flow of signals between layers corresponding to different levels of abstraction. The diagram is labelled with suggestive interpretations of the various stages and processes in this operation, but

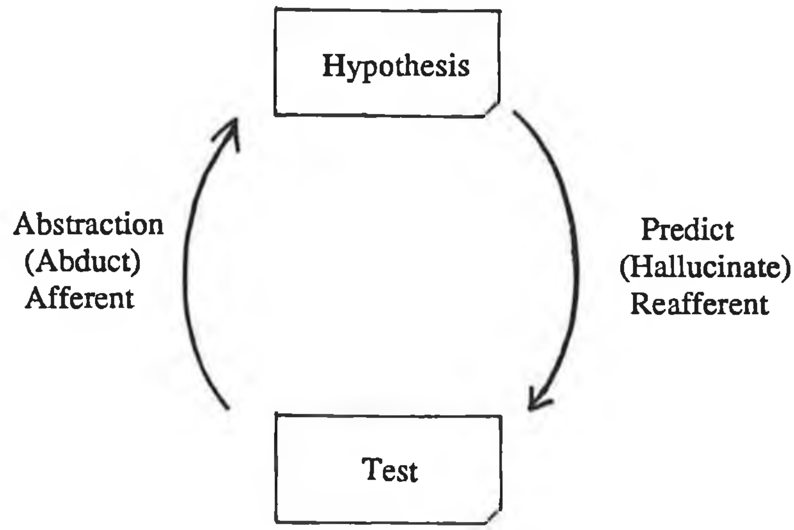


Figure 35. Schematic illustration of the basic signal flow in a sensory information processing primitive.

these are intended only for explanatory purposes. The operational description of this loop would be purely in terms of signals (i.e., information theoretic concepts) and/or dynamics. In this picture, the process of making a decision might be interpreted as a back and forth flow of signals until a resonant state is arrived at which involves consistent or compatible "interpretations" between the different levels. In this sense, the level at which a decision-making or recognition process takes place, even though it can be formulated in information theoretic terms, is a different level of description from the processes that control the connections of signals into and out of this system. There are parallels between this picture and the work of Freeman on the olfactory system (the sense of smell). He describes recognition as chaotic attractor states of coupled systems of distributed oscillators. Some initial work by Marshall on coupled oscillators used for recognition also supports these conclusions.

However, these suggestions about the signal or information processing mechanisms involved in perception, are not in themselves the conclusions of this research. The original aim described above was to ask what we should be doing in order to make progress in vision. The conclusion is that we should not be trying to define how to map from scenes or images to 3-D representations, say. We should be trying to understand the information or signal processing properties of certain types of dynamics. We should be trying to understand and model the dynamics of perceptual systems in biological organisms. We should be trying to formulate a more complete picture which encompasses the results of self-organization-based information processing ideas, with the results on recognition based on relative information and signal-to-symbol transitions.

To make this more concrete, consider for instance Grossberg's model of the early stages of human visual perception, which is a phenomenological model, formulated in terms of differential equations. The model seems to be uniquely successful at describing the properties of certain aspects of human visual perception like object recognition, intensity, colour, texture and stereo perception, and so on. The fact that this is a phenomenological model means that it gives little clue of the developmental pressures which might have led to this particular type of processing. But these questions about development are on the other hand, the very types of questions answered by Linsker's work. Unfortunately, at the moment there is not a unified way of coupling the domains of these two sets of research results. The fact that Grossberg's models are dynamic models, described in terms of differential equations, gives no clue about whether the processes can be described in other terms, such as algorithmic terms that do not need the painstaking incremental integration of the differential systems. Reinterpreting Grossberg's results for example, in information theoretic terms which are compatible with other results, like Linsker's, already expressed in informational terms would give tremendous insight into the mechanisms of perception. The conclusion of this dissertation is that this is the sort of thing we need to do. A theoretical framework for doing this, centred around information theory is presented. The mathematical formalism based on this framework, within which explanatory models of perceptual capacities could be constructed, is the next step in this research process.

Glossary

Accommodation Adjustment of the optics of an eye to keep an object in focus on the retina as its distance from the eye varies. In the human eye this is achieved by varying the thickness of the lens.

Action potential The interior of a nerve cell has a negative electrical charge relative to the exterior. If the axon of a nerve cell is stimulated electrically, the membrane allows current to cross it and the charge is momentarily reversed. This change in membrane behaviour spreads rapidly down the axon and the wave of change in voltage across the membrane which it causes is called an action potential.

Activator In a dynamical system with many interacting components, one component is an activator of another if it tends to increase the activity (rate of production) of the other.

Agonist In a dynamical system with many interacting components, a component is an agonist for another, if it tends to increase the effect of the other on the activity (rate of production) of a third.

Aliasing The generation of low frequency artifacts when a high frequency signal is sampled below the Nyquist frequency (twice the maximum frequency contained in the signal).

Allonomy Literally external law. Used by Maturana and Varela to refer to control, or input-process-output type systems where the system is *defined* in terms of its input and output by some external agent.

Amacrine cell A type of cell in the vertebrate retina.

ANN Synonym for artificial neural network

Antagonist In a dynamical system with many interacting components, one component is an antagonist for another if it tends to decrease the affect the other has on the activity (rate of production) of a third.

Anti-realism In the philosophical debate on the status of universals, it is the view that universals are not real and are not needed for the perception of particulars (objects).

Autonomy Literally self-law. Used by Maturana and Varela to refer to systems with a circular organisation which are self-defining or assert their own identity. They can be perturbed by external influenced but are not defined by these. A more general concept than autopoiesis below.

Autopoiesis Literally means self-production. An *autopoietic* system is organised (defined as a unity) as a network of processes of production (transformation and destruction) of components that: (1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them: and (2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network. It is intended by Maturana & Varela as a definition of the minimum criterion for something to be *living*.

Axiology The theory of value

Backpropagation In the field of artificial neural networks, the backpropagation rule is used to feed the output error, (difference between actual output and desired output) back to affect the weights of connections between previous layers in a multi-layer network, in a way which is proportional to the effect they have on this output.

Bandwidth The range of signal frequencies over which a signal processing device can operate.

Bausteinentropie Literally building block entropy. It is simultaneously a measure of the individual entropies of parts of a multi-partite system and of the entropy of the whole system. Introduced by Watanabe, it is related to the notion of redundancy in information theory.

Bayesian inference Statistical inference based on Bayes theorem.

Bipolar cell A type of cell in the vertebrate retina.

Boolean algebra An algebraic system consisting of a set of elements, together with two binary operations obeying certain axioms. Used as an algebraic formulation of logic. Effectively equivalent to mathematical system of set theory.

Borel fields are effectively equivalent to sets of subsets of a given set, but can involve countably infinite unions and intersections of these subsets.

Broadband or wideband A frequency band that extends over a wide range of frequencies.

Complex system Term used by Robert Rosen to describe systems which cannot be accurately modelled by a Newtonian type of dynamical system.

Complex cell Cell in the visual cortex responding either to an edge, a bar or a slit stimulus of a particular orientation falling anywhere within its receptive field.

Compound eye An eye constructed of many *ommatidia*, each one a small elongated eye-cup with a crystalline cone at its tip and the light sensitive *rhabdom* below it. There are two main types: the *apposition* type where light passing through a particular cone lens is mostly absorbed by the receptor of that ommatidia; the *superposition* type where there is a clear space between the cones and rhabdoms making more efficient use of light.

Computational theoretic A term introduced by Marr. Computational theories of vision are concerned with how, in principle, particular kinds of information such as the shapes of objects or distances of surfaces can be extracted from images. Solutions to such problems involve consideration of the constraints which apply to the structures of natural objects and surfaces and the ways in which they reflect light. An example is the demonstration that the shape of an object can be recovered from its silhouette if the shape approximates to a generalised cone.

Cone receptors Cone shaped photoreceptors in the vertebrate retina which are sensitive to the wavelength of light in normal lighting conditions.

Connectionism Research program within cognitive science concerned with the properties of dynamical systems consisting of very simple interconnected *nodes*. The connections, which carry the output activity value of one node one of the many inputs of another node, often have variable weights affecting the strength of the interconnection. See also artificial neural networks, PDP, emergence.

Control theory An extension of dynamic systems theory which puts the emphasis on forcing a system to follow or tend to a particular trajectory under the influence of externally determined inputs. See also allonomy.

Cortical magnification factor A factor describing the extent to which a particular area on a certain part of the retina, is mapped on to an area of a particular size on the cortex. This factor changes substantially across the retina from high values for the fovea (represented by a large area of cortex) to low values in the far periphery (represented by small areas in the cortex).

Deductive inference An inference where the conclusion are evidentially implied by the premises.

Dendrites The processes of nerve cells which carry slow potentials from synapses to the cell body.

Directional selectivity A difference in the response of a cell to a pattern of light moving through its receptive field according to the direction of movement.

Doghood The property of being, or belonging to the class of, a dog.

Edge-detector Strictly, a detector for "edges" or intensity changes in an image which are supposed to correspond to object edges or contours in the real world.

Eigenimages Eigenvector of a matric which is the covariance matrix of an ensemble of images represented as vectors.

Entropy Roughly speaking, the entropy of a probability distribution measures how chaotic the distribution is. Thus the entropy is low if the distribution is concentrated around one value, and zero if it is concentrated on exactly one value. A uniform distribution has the maximum entropy. Formally, for a discrete distribution with outcomes 1,2,...,i,... having probabilities $p_1, p_2, \ldots, p_i, \ldots$, the entropy $q(p)$ of the distribution is given by $q(p) = \sum_i -p_i \log p_i$

Ergodicity A topic dealing with the relationship between statistical averages and sample averages.

Essentialism A term Popper uses to denote the classical realist position on the status of the universal.

Exact differential A differential form which is the total derivative of some function. That is an exact differential can be integrated.

Extension The collection of all objects corresponding to a concept or satisfying a predicate. Compare *intension.*

Extra-evidential Without the support of evidence.

Extra-logical Not the basis of a logic deduction or inference.

Extra-fovea The part of the retina immediately outside of and surrounding the fovea (between 1° and 5-10° off the visual axis).

Fovea Pit-shaped depression in a vertebrate retina, usually in an area centralis.

Frege principle A term used by Watanabe to indicate the view that every predicate has a finite extension, i.e., a finite and definite set of objects in the set corresponding to that predicate. It is a notion closely aligned with Boolean logic and set theory.

Gabor elementary function A Gaussian modulated sinusoid signal in the time or space domain. It minimizes the product of spread in both time (or space) and frequency domains. It is equivalent to the minimum uncertainty function in quantum mechanics.

Ganglion cell A type of cell in the vertebrate retina. The axons of ganglion cells are packed together in the optic nerve and carry information from retina to brain.

Gap junctions Interactions between neural processes (dendrites, axons), involving purely electric effects rather than the more usual chemical transmission that takes place at synapses.

GEF Acronym for Gabor elementary function

Graded potential or slow potential. A potential difference between the inside and outside of a neural membrane which is relatively constant with time (compared with a "spiking potential"). They do not travel far along dendrites without significant attenuation.

Holonomic A kinematic condition on a dynamic system that allows the elimination of a coordinate variable and a velocity (or momentum) variable, thus reducing the number of configurational degrees of freedom of the system by one and the number of dimensions of the phase space by two. It corresponds to a differential form in differentials of the coordinates (or velocities) which is exact or integrable.

Homunculus Literally a tiny man. Associated with the fallacy that in order for us to see there must be a little man in our head looking at images coming from the eyes.

Horizontal cell A type of cell in the vertebrate retina.

Hyperacuity Humans can carry out a variety of tasks to accuracies that are more precise than the dimensions of the retinal cones from which the information originates. Foveal cones have a diameter of about 27" (of arc), yet many tasks yield accuracies of around 5", and stereoscopic acuity may be as good as 2". Such tasks are said to fall within the range of hyperacuity.

Hypercolumn A block of the visual cortex in which all cells have receptive fields falling in a single area of the retina.

Hyperpolarisation A change in the membrane potential of a nerve cell such that the interior becomes more negatively charged relative to the exterior. If the membrane of an axon is hyperpolarised, action potentials are generated with lower frequency.

Imagelets A term used to indicate that the objects of interest are images of a sort, but with far fewer pixels than would typically be the case.

Inductive inference An inference for which there is insufficient evidence to necessarily follow from the premises.

Information theory A theory of the properties of signal sources and communication channels based on the average or statistical properties of these. Associated with the name of Shannon.

Inhibitor In a dynamical system with many interacting components, one component is an inhibitor of another if it tends to decrease the activity (rate of production) of the other.

Inner plexiform layer Layer in the vertebrate retina where the bipolar, ganglion and amacrine cells synapse with each other.

Intension The collection of predicates just sufficient to describe a concept.

Isopreference curves Curves drawn on a graph of different physical parameters indicating that subjects had roughly equal preference for say, images, corresponding to a particular curve, despite the variation of parameters along the curve.

Karhunen-Loève transform A mathematical transform based on the transformation from a default coordinate system to one where the axes are the eigenvectors of the covariance matric of the original data.

KLT Synonym for the Karhunen-Loève transform.

Kolmogorov Responsible for the axiomatic approach to probability theory

Large monopolar cells Neural cells in the compound eye of the fly which are roughly the anatomical analogue of the bipolar cells in the vertebrate retina.

Lateral geniculate nucleus (LGN) The part of the mammalian brain where the axons of retinal ganglion cells terminate, and from which axons run to the visual cortex.

Logon A term used by Gabor to denote the "minimum uncertainty" signal which is now usually called the Gabor elementary function.

Mach band phenomenon A subjective perception of light or dark bands on either side of the boundary where two regions with different reflectance meet.

Metadynamics A term used mostly within the Alife research community to describe adaptation which actually involves change in the dynamic structure of a system.

Moiré fringes The low spatial frequency interference pattern formed when we look through two overlaid grids or gratings with high spatial frequency.

Modulation Transfer Function (MTF) The amplitude of the Fourier transform of a filter or function. The MTF is useful because by looking at its graph, one can tell at a glance which frequencies are passed and which are suppressed by the filter.

Newtonian paradigm The use of Newtonian type dynamical systems to describe a system which does not necessarily involve the interaction of physical "particles".

Nominalism A position opposed to realism which holds that there are no general kinds like doghood, only particular words.

Non-holonomic constraint A kinematic condition on a dynamical system which allows the elimination of a velocity variable or a differential of a configurational variable, but does not allow the corresponding configurational variable itself to be eliminated. Thus the freedom to vary a particular velocity is removed but not the freedom to vary the corresponding configurational variable and in other words without reducing the number of degrees of freedom. It corresponds to a differential form which expresses a connection between the velocities (or coordinate differentials) but cannot be integrated to give a related constraint on the configurational coordinates themselves.

Nyquist limit The highest frequency that can be sampled in a digital system without causing aliasing.

Octave An increase in frequency of one octave is a doubling of frequency.

Ommatidium Unit of the compound eye containing a light-sensitive rhabdom.

Opponent-colour response If light of one wavelength falling in the receptive field of a cell causes that cell to fire more frequently than its resting rate, and light of a different wavelength causes it to fire less frequently, the cell is said to have an opponent-colour response.

Outer plexiform layer A layer in the vertebrate retina where receptors, bipolar and horizontal cells all synapse with each other.

Paradigm Either the original sense of (i) a pattern, example, model, class sample, or (ii) the Kuhnian sense of an ideological theory or approach to scientific problems.

Paradigmatic symbol A term introduced to Watanabe to describe what human thought operates with, as opposed to an abstract symbol which a computer operates with. It is used in th sense of a particular example "standing for" or "eliciting" all other examples of a class.

Periphery The part of the retina greater than approximately 10° off the visual axis.

Phase space For a dynamic system model with N degrees of freedom, the phase space is the 2N-dimensional space with the configurational variables, and their velocities (or momentum) as coordinates. The state of the system is uniquely defined by giving its position in phase space (or its position and velocity in configuration space, for comparison).

Phenomenology Literally, the description or study of appearances. Any description of how things appear, especially if sustained and penetrating.

Phenomenological domain is defined by the properties of the unity or unities that constitute it, either singly or collectively through their transformations or interactions. Thus, whenever a unity is defined, or a class or classes of unities are established that can undergo transformations or interactions, a *phenomenological domain* is defined [24, p.46].

Photoreceptor A receptor cell sensitive to light.

Platonist In mathematics, Platonists, or realists, think that abstract concepts, like numbers, are entities, and that mathematical truth, including those about infinite numbers, exist independently of our researches. Compare formalist and constructivist.

Point spread function The 2-D spatial equivalent of the *impulse response* in a time-varying system in engineering, or a *Green function* in the mathematics of differential equations, or the *propagator* in physics, or the *kernel* in quantum mechanics.

Polychotomy A choice between several mutually exclusive possibilities.

Predicate What can be said of (predicated of) a subject. In 'Grass is green', *grass* is the subject and *green* is the predicate.

Processes One sense of the term refers to the elongated extensions of nerve cells divided into dendrites and axons.

Psychophysics The analysis of perceptual processes by studying the effect on a subject's experience or behaviour of systematically varying the properties of a stimulus along one or more physical dimensions.

Raw primal sketch In Marr's theory of vision, a rich representation of the intensity changes present in the original image.

Reafferent *Afferent* designates a nerve cell that transmits ingoing signals from the peripheral receptors to the central nervous system. *Reafferent* designates signals that go from the CNS back towards the peripheral sensory receptors. Distinguish *efferent*.

Receptive field The area of the retina in which light causes a response in a particular nerve cell.

Refractive index A measure of the extent to which a medium refracts light.

Retinal eccentricity Angular distance of a point on the retina from the centre of the fovea.

Retinotopic map An array of nerve cells which have the same positions relative to one another as their receptive fields have on the surface on the retina.

Rod receptors Vertebrate photoreceptor with long outer segment, sensitive to light of low intensity.

Saccade Rapid movement of the eye to fixate a target.

Simple cell Cell in the visual cortex showing linear spatial summation of light intensities in parts of its receptive field separated by straight line boundaries.

Simple systems Used by Rosen to designate systems that can be fully modelled in terms of the rate equations of a Newtonian dynamics.

Solipsism Literally, 'only-oneselfism' The view that nothing exists outside one's own mind, or that nothing such can be known to exist.

Soma The cell body of a neuron.

Spiking potential or action potential A relatively large fluctuation in the potential across the membrane of a neural axon from the normal resting potential of -60mV to a positive potential of 20-30mV and back in <1ms. Usually they appear as a train of pulses travelling down the axon. They can travel large distances (1-2m) along axons at high speed without significant attenuation.

Stereopsis Perception of depth dependent upon disparity in the images projected on the retinas of the two eyes.

Striate cortex The primary visual cortical receiving area in the monkey and in man. So called because of the stria of Genarii, a band of white matter running through only this region of the cortex.

SVD Synonym for singular value decomposition

Synapse A point where the membranes of two nerve cells nearly touch and where electrical activity in one cell influences the membrane potential of the other cell.

Topographic map A map which preserves in its range, the spatial ordering of its domain.

Transduction The process by which external energy impinging on a receptor cell causes a change in its membrane potential.

Visual cortex Primary region of the mammalian cortex in the occipital lobe receiving input from the LGN. Cells in the primary visual cortex respond to light falling on the retina and are arranged in a retinotopic map. In primates, this region is also known as the striate cortex[5].

---

[5]References [9] and [56] are gratefully acknowledged for assistance in preparing this glossary.

References

1. Mooney,G. & Murphy,N., *Three-Dimensional Computer Vision for Robotic Assembly - a Preliminary Report*, TR01/1985, School of Electronic Engineering, NIHE, Dublin.

2. Brown,J., O'Kelly,M.E.J. & Stewart,I., *Robotics and Automation in Batch Manufacture*, NBST Report, July 1983.

3. Batchelor, B.G., *Intelligent Image Processing in Prolog*, Springer-Verlag, London 1991.

4. Power, L. *Vision in Industrial Inspection*, M.Eng. Thesis, Dublin City University, in preparation.

5. Power,L.S., Automated Measurement of Variation in Manufacture, *Proceedings of the Eight Conference of the Irish Manufacturing Committee (IMC-8)*, University of Ulster at Jordanstown, September 1991, pp.701-714.

6. Batchelor,B.G., Tools for Designing Industrial Vision Systems, *Proceedings of the DCU 1992 Computer Vision Seminars*, DCU School of Electronic Engineering Technical Report, 1992 (to be published).

7. McClannon, K.M., Whelan, P.F. & McCorkell, C., Machine Vision in Process Control, 'Closing The Loop', in *Proceedings of the Ninth Conference of the Irish Manufacturing Committee, IMC-9*, University College Dublin, September 1992.

8. Whelan,P.F. & Batchelor,B.G., *Development of a vision system for the flexible packing of random shapes*, Proc. SPIE Machine Vision Applications, Architectures and Systems Integration Conference, Vol.1823, Boston, 14-20 Nov. 1992, (to be presented).

9. Marr, D., *Vision: A computational investigation into the human representation and processing* of visual information, *W.H.Freeman & Co., San Francisco, 1982.*

10. Murphy, N. *Update report on NBST funded project.* Presentation at annual NBST project review meeting. Not published.

11. Levine, M.D., *Vision in Man and Machine*, McGraw-Hill, 1985.

12. Rosen, R., 1985, Organisms as Causal Systems which are not Mechanisms: An essay into the Nature of Complexity, in *Theoretical Biology and Complexity*, Rosen, R. (ed.), Academic Press. See also Chapter 7 (Appendix) of Rosen, R., 1985, *Anticipatory Systems*, Pergamon Press.

13. Livingstone, M. and Hubel, D., Segregation of Form, Colour, Movement and Depth: Anatomy, Physiology and Perception, *Science*, Vol. 240, 6 May 1988, pp.740-749.

14. Livingstone, M., Art, Illusion and the Visual System, *Scientific American*, Vol.258, pp.68-75, Jan. 1988.

15. Laughlin, S.B., "Form and function in retinal processing", *Trends In Neuroscience (TINS)*, Vol.10, No.11, 1987, pp.478-483.

16. Barlow, H.B., "Critical limiting factors in the design of the eye and visual cortex – The Ferrier Lecture 1980", *Proc. Royal Society London Series B*, Vol.212, 1981, pp.1-34.

17. Barlow, H., and Földiák, P, Adaptation and Decorrelation in the Cortex, pp.54-72 in Durbin, R.M., Miall, R.C., & Mitchison, G.J. (eds.), *The Computing Neuron*, Addison-Wesley, 1989.

18. Linsker, R., Self-Organization in a Perceptual Network, *IEEE Computer Magazine*, March 1988, pp.105-117.

19. Watanabe, S., *Pattern Recognition - Human and Mechanical*, Wiley, New York, 1985.

20. Rosen,R., *Fundamentals of Measuerment and Representation of Natural Systems*, North-Holland, NY, 1978.

21. Rosen,R., *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations*, Pergamon Press, 1985.

22. Dretske,F.I., *Knowledge and the Flow of Information*, MIT Press/Bradford, 1981.

23. Maturana,H.R. & Varela,F.J., *Autopoiesis and Cognition: The Realization of the Living*, Boston Studies Series in the Philosophy of Science, D. Reidel, Holland, 1979.

24. Varela, F.J., 1979, *Principles of biological autonomy*, Elsevier North Holland, NY.

25. Maturana H.R. & Varela F.J., 1987, *The Tree of Knowledge, The Biological Roots of Human Understanding*, Shambhala, Boston.

26. Varela,F.J., Autopoiesis and a Biology of Intentionality, in *Proceedings of ESPRIT BRA workshop on Autopoiesis and Perception*, McMullin,B. & Murphy,N., (eds.), held in Dublin City University, August 1992, to be published.

27. Sandini,G. (ed.), *Computer Vision - Proceedings of the Second European Conference on Computer Vision (ECCV-2)*, S. Margherita Ligure, Italy, May 1992, Springer-Verlag 1992.

28. Cariani,P., *On the design of devices with emergent semantic functions*, Ph.D. thesis, State University of New York at Binghamton, 1989.

29. Boden M., 1988, *Computer Models of the Mind*, Cambridge University Press, UK.

30. Varela, F.J., 1992, Whence perceptual meaning? A cartography of current ideas, pp.235-263, in F.J. Varela and Jean-Pierre Dupuy (eds.), *Understanding Origins*, Kluwer, The Netherlands.

31. Hinton,G., Learning in parallel networks, *Byte*, No.10, p.265, 1985.

32. Rummelhart,D., McClelland,J. and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol I: Foundations,* MIT Press, 1986.

33. Hubel, D.H. & Wiesel, T.N., 1977, Functional architecture of macaque monkey visual cortex, *Proceedings of the Royal Society of London, series B,* Vol 198, pp1-59.

34. Barlow, H.B., 1972, Single units and sensation: a neuron doctrine for perceptual psychology?, *Perception*, Vol 1, pp371-94.

35. Grey, C.M., Konig, P. Engel, A.K. & Singer, W., 1989, *Nature*, Vol 338, pp334-337.

36. Eckhorn, R. *et al*, 1988, *Biological Cybernetics*, Vol 60, pp121-130.

37. Freeman, W.J., 1991, The Physiology of Perception, *Scientific American*, pp34-41, February.

38. Kohonen,T., *Self-Organization and Associative Memory*, 2nd ed., Springer-Verlag, 1988.

39. Skarda,C.A., Perception, Connectionism and Cognitive Science, pp.265-271, in F.J. Varela and Jean-Pierre Dupuy (eds.), *Understanding Origins*, Kluwer, The Netherlands, 1992.

40. Freeman,W.J., The Physiology of Perception, *Scientific American*, Feb. 1991, pp.34-41.

41. Zak,M., An Unpredictable-Dynamics Approach to Neural Intelligence, *IEEE Expert Magazine*, Aug. 1991, pp.4-10.

42. Shannon, C.E., "Communication in the presence of noise", *Proc. IRE*, Vol.37,pp.10-21, June 1949; Reprinted in *Proc. IEEE*, Vol.72, No.9, Sept. 1984, pp.1192-1201.

43. Lacey, A. R., *A Dictionary of Philosophy*, 2nd ed., Routledge & Kegan Paul, London, 1986.

44. Popper, K., *Unended Quest – An Intellectual Autobiography*, Flamingo,. 1986.

45. Murphy,N., The Causal and Symbolic Explanatory Duality as a Framework for Understanding Vision, in *Proceedings of ESPRIT BRA workshop on Autopoiesis and Perception*, McMullin,B. & Murphy,N., (eds.), held in Dublin City University, August 1992, to be published.

46. Jammer, M., *The Philosophy of Quantum Mechanics: The Interpretations of Quantum Mechanics in Historical Perspective*, Wiley, 1974.

47. Watanabe, S., *Knowing and Guessing – A quantitative study of inference and information*, Wiley, New York, 1969.

48. Von Neumann, J., *The Mathematical Foundations of Quantum Mechanics*, (trans. R.T. Beyer), Princeton, Princeton University Press, 1955.

49. Birkhoff, G. and von Neumann, J., The Logic of Quantum Mechanics, *Annals of Mathematics*, Vol. 37, No.4, October 1936, pp.823-843.

50. Zukav, G. *The Dancing Wu Li Masters*, Flamingo, 1989. (1st pub. W .Morrow, 1979).

51. Bunge, M., *Causality and Modern Science*, 3rd ed., Dover, NY, 1979.

52. Goldstein, H., *Classical Mechanics*, 2nd ed., Addison-Wesley, 1980.

53. Varela, F.J. & Bourgine, P., (eds.), Towards a practice of autonomous systems, Introduction to *Proceedings of the first European Conference on Artificial Life, ECAL1*, Paris 1991, MIT Press, 1992.

54. Bennett, B.M., Hoffman, D.D., Prakash,C., *Observer Mechanics – A Formal Theory Of Perception*, Academic Press, 1990, p.8.

55. Peirce, C.S., Collected Papers, Volume II, Harvard University Press, Cambridge, Mass., 1931-35. See also [19, p.101].

56. Bruce, V., and Green, P.R., *Visual Perception – Physiology, Psychology and Ecology*, 1st edition, 1985, Lawrence Erlbaum Associates, pp.31-32.

57. Crick, F.H.C., *"Thinking about the Brain"*, *Scientific American*, Vol.241, No.3, p219, 1979.

58. Piccolino, M., "Cajal and the retina: a 100-year retrospective", *Trends in Neuroscience*, Vol.II, No. 12, 1988, pp.521-525.

59. Marr, D. and Hildreth, E., A theory of edge detection, *Proc R. Soc. Lond. Series B*, Vol.207, pp.187-217.

60. Pratt, W.K., *Digital Image Processing*, Wiley, 1987.

61. Hubel, D.H., *Eye, Brain and Vision*, Scientific American Library Series; No.22, Freeman, 1988.

62. Williams, D.R., "Seeing through the photoreceptor mosaic", *Trends in Neuroscience*, May 1986, pp.193-198.

63. Snyder, A.W., and Miller, W.H.,"Photoreceptor diameter and spacing for highest resolving power", *J. Opt. Soc. Am.*, 67, 1977, pp696-698.

64. Cornsweet, T.N., *Visual Perception*, Academic Press, NY, 1970.

65. Nilsson, D., "From Cornea to retinal image in invertebrate eyes", *Trends in Neuroscience*, Vol. 13, No. 2, 1990, pp55-64.

66. Sur,M., Pallas,S.L. & Roc,A.W., Cross-modal plasticity in cortical development: differentiation and development in the sensory neocortex, *Trends in Neuroscience*, Vol.13, No.6, 1990, pp.227-233.

67. Williams, D.R., *J. Opt. Soc. Am.*, Vol. A.2., 1985, pp1087-1093.

68. Williams, D.R., & Collier, R., "Consequences of Spatial Sampling by a Human Photoreceptor Mosaic, *Science*, Vol. 221, 22 July 1983, pp385-387.

69. Yellot, J.I.Jr., "Spectral Consequences of Photoreceptor Sampling in the Rhesus Retina", in *Science*, Vol. 221, 22 July, 1983, pp382-385.

70. Bossomaier, T.R.J., Snyder, A.W., & Hughes, A., *Vision Research*, Vol.25, 1985, pp145-147.

71. Field, D.J., "Relations between the statistics of natural images and the response properties of cortical cells", in *J. Opt. Soc. Am.*, Vol. A4, No. 12, Dec. 1987, pp2379-2394.

72. Kuffler, S.W., "Discharge patterns and functional organisation of mammalian retina", *Journal of Neurophysiology*, 16, 1953, pp37-68.

73. Enroth-Cugell, C., & Robson, J.G., "The contrast sensitivity of retinal ganglion cells of the cat", in *Journal of Physiology*, Vol.187, 1986, pp517-552.

74. Hubel, D.H., & Wiesel, T.N., "Receptive fields and functional architecture of monkey striate cortex", in *Journal of Physiology*, Vol.195, 1968, pp215-243.

75. McIlwain, J.T., "Point images in the visual system: new interest in an old idea", *Trends in Neuroscience*, Aug. 1986, pp354-358.

76. Lemon, R., "Cognitive control of movement (News and Views)", *Nature*, Vol.337, 2 Feb. 1989, p.410.

77. Sejnowski, T.J., "Neural populations revealed (News and Views)", *Nature*, Vol.332, 24 March 1988, p.308.

78. Georgopoulos, A.P., et.al., "Mental rotation of the Neuronal Population Vector", *Science*, Vol.243, 13 Jan. 1989, pp.234-236.

79. Fischer, B., *Vision Research*, Vol.12, 1972, pp.1125-1144.

80. Fischer, B., *Vision Research*, Vol.13, 1973, pp.2113-2120.

81. Peichl, L.,and Wassle, H., *J. Physiol.* (London), Vol.291, 1979, pp.117-141.

82. Albus, K., *Exp. Brain Res.*, Vol.24, 1975, pp.159-179.

83. Hubel, D.H.,and Wiesel, T.N., *J. Comp. Neurol.* Vol.158, 1974, pp.295-306.

84. Hubel, D.H., and Wiesel, T.N., *Proc. R. Soc. London Series B*, Vol.198, 1977, pp.1-59.

85. Dow, B.M., and Snyder, A.Z., Vautin, R.G., and Bauer, R., *Exp. Brain. Res.*, Vol.44, 1981, pp.213-228.

86. Van Essen, D.C., Newsore, W.T., and Maunsell, J.H.R., *Vision Res.*, Vol.24, 1984, pp.429-448.

87. Wassle, H., Grunert, U., Rohrenbeck, J., and Boycott, B.B., *Nature*, Vol.341, 19 Oct. 1989, pp.643-646.

88. Hinton, G., in *Proc. 7th Intl. Joint Conf. on AI*, Vol.2., Kaufman, 1981, pp.1088-1096.

89. Sterling, P., Freed, M., and Smith, R.G., "Microcircuitry and functional architecture of the cat retina", Trends In Neuroscience, May 1986, pp.187-192.

90. Sterling, P., and Kaneko, A., "Structure – Function Relationships in the Vertebrate Retina – An Introduction", Trends In Neuroscience, May 1989, Centrefold.

91. Nelson, R., and Kolb, H., *Vision Research*, Vol.23, 1983, pp.1183-1195.

92. Masland, R.H., "The Functional Architecture of the Retina", *Scientific American*, pp.90-99, Vol.255, p.90, 1986.

93. Masland, R.H., "Amacrine Cells", *Trends In Neuroscience*, Vol.11, No.9, 1988, pp.405-410.

94. Masland, R.H., and Tauchi, M., "The Cholinergic Amacrine Cell", *Trends In Neuroscience*, May 1986, pp.218-223.

95. Dowling, J.E., "Dopamine: a retinal neuromodulator", *Trends In Neuroscience*, May 1986, pp.236-240.

96. Koch, C., Poggio, T.,& Torre, V., "Computations in the vertebrate retina: gain enhancement, differentiation and motion discrimination", *Trends In Neuroscience*, May 1986, pp.204-211.

97. Miller, R.F., "Are single retinal neurons both excitatory and inhibitory, (News and Views)," *Nature*, Vol.336, 8 Dec. 1988, pp.517-518.

98. Shapely, R., and Perry, V.H., "Cat and Monkey retinal ganglion cells and their visual functional roles", *Trends In Neuroscience*, May 1986, pp.229-235.

99. Srinivasan, M.V., Laughlin, S.B., and Dubs, A., (1982), *Proc. R. Soc. London Series B*, Vol.216, pp.427-459.

100. Field, D.J., "Relations between the statistics of natural images and the response properties of cortical cells", *J. Opt. Soc. Am. A*, Vol. 4, No. 12, Dec. 1987, pp.2379-2394.

101. Buchsbaum, G., and Gottschalk, A., *Proc. R. Soc. London, Series B*. Vol.220, 1983, pp.89-113.

102. Buchsbaum, G., *Proc. IEEE*, Vol.69, No. 7, July 1981, pp.772-786.

103. Barlow, H.B., "The coding of sensory messages", in *Current Problems in Animal Behaviour*, W. H. Thorpe & Zangwill, O.L., (eds.), Cambridge University Press, Cambridge, 1961, pp.331-360.

104. Barlow, H.B., "Three theories of cortical function", in *Development Neurobiology of Vision*, R.D. Freeman, ed., (Plenum, N.Y.), 1979, pp.1-16.

105. Barlow, H.B., "Understanding natural vision", in *Physical and Biological Processing of Images*, O.J. Braddick and A.C. Sleigh, eds., Vol II of Springer Series in Information Sciences (Springer-Verlag, Berlin), 1983, pp.2-14.

106. Torre,V. & Poggio,T., On edge detection, *IEEE T.PAMI*, Vol.8, No.2, 1986, pp.147-163.

107. Spoehr, K.T., and Lehmkuhle, S.W., *Visual Information Processing*, Freeman, 1982.

108. Collins Pocket Dictionary of the English Language, Collins, 1981.

109. Boden, M.A., *Computer models of mind*, Cambridge University Press, 1988.

110. Fischler, M.A., and Firschein, O., *Intelligence: the eye, the brain and the computer.*Addison-Wesley, 1987.

111. Best, J.B., *Cognitive Psychology*, West, 1986.

112. Perkel, D.H., "Logical Neurons: the enigmatic legacy of Warren McCulloch, Trends In Neuroscience, Vol.11, No.1, 1988.

113. Judson, H.F., *The Search for Solutions*, Hutchinson, 1980.

114. Daintith,J., (ed.), *A Dictionary of Physical Science*, Pan 1978.

115. Tolman, R.C., *The Principles of Statistical Mechanics*, Dover, 1979. (First published by O.U.P. in 1938.)

116. See Ref. 14, p.162 for appropriate references. Also Szilard, *L.Z.Phys*, Vol.53, 1929, p.840.

117. Watanabe, S., "Uber die Anwendung Thermodynamischer des Atomkerns, *Zeitschrift fur Physik*, Vol.113, 1939, (in German), p482.

118. Watanabe, S., "Correlation Phenomena in Atomic Nuclei, Parts I, II, III, *Kagaku*, Vol.14, 1944, (in Japanese), p.82, p.122, p.168.

119. Nyquist, H., "Certain factors affecting telegraph speed", *Bell System Technical Journal*, April 1924, p.324.

120. Hartley, R.V.L., "Transmission of Information", *Bell System Technical Journal*, July 1928, p.535.

121. Shannon, C.E., and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1949.

122. Shannon, C.E., "Communication in the presence of noise, *Proc. IRE*, Vol.37, pp.10-21, June 1949. Reprinted in *Proc. IEEE*, Vol.72, No.9, Sept. 1984, pp.1192-1201.

123. Papoulis, A., *Probability, Random Variables and Stochastic Processes*, McGraw–Hill, 1984.

124. Slepian, D., "On Bandwidth", *Proc. IEEE*, Vol.64, No.3, March 1976, pp.292-300.

125. Slepian, D., and Pollak, H.O., "Prolate spheroidal wave functions; Fourier analysis and uncertainty – part I", *Bell Syst. Tech. J.*, Vol.40, 1961, pp.43-64.

126. Landau, H.J., and Pollak, H.O., "Prolate spheroidal wave functions; Fourier analysis and uncertainty – parts II and III", *Ibid*, Vol.40, 1961, pp.65-84; Vol.41, 1962, pp.1295-1336.

127. Nyquist, H., "Certain topics in telegraph transmission theory", *A.I.E.E. Trans.*, April 1928, p.617.

128. Gabor, D., "Theory of Communication", *Proc. Inst. Elec. Eng.*, Vol.93, No.26, 1946, pp.429-441.

129. Ville, J.,"Theorie et applications de la notion de signal analytique", *Cable et Transmission*, Vol.2, No.1, 1948, pp.61-74.

130. Page, C.H., Instantaneous Power Spectra, *J. Appl. Phys.*, Vol.23, pp.102-106, 1952.

131. Cohen, L., "Time Frequency Distribution – A Review", *Proc. IEEE*, Vol.77, No.7, July 1989.

132. Porat, M., and Zeevi, Y.Y., "The Generalised Gabor Scheme of Image Representation in Biological and Machine Vision", *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol.10, No.4, July 1988.

133. Wigner, E., "On the quantum correction for thermodynamic equilibrium", *Phys. Rev.*, Vol.40, June 1932, pp.749-759.

134. Daugman, J.G., "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Am. A*, Vol.2, No.7, pp.1160-1169.

135. Reed, T.R., and Wechsler, H., "Segmentation of Textured Images and Gestalt organisation using spatial/spatial frequency representations", *IEEE T.PAMI*, Vol.12, No.1, Jan.1990.

136. Jacobson, L.,and Wechsler, H., "Joint spatial/spatial-frequency representation", *Signal Processing*, Vol.14, No.1, Jan.1988, pp.37-68.

137. Cohen, L., "Generalised phase-space distribution functions", *J. Math. Phys.*, Vol.7, No.5, May 1966, pp.781-786.

138. Huang, T.S., "PCM Picture Transmission", *IEEE Spectrum*, Vol.2, No.12, pp.57-63.

139. Gonzalez, R.C. & Wintz, P.A., *Digital Image Processing*, Addison-Wesley, 1978.

140. Davidson, M., "Perturbation Approach to Spatial Brightness Interaction in Human Vision", *J. Opt. Soc. Am.* , Vol.58, No.9, Sept.1968, pp.1300-1308.

141. Stockham, T.G., "Image Processing in the context of a visual model", *Proc. IEEE*, Vol.60, NO.7, July 1972, pp.828-842.

142. Hall, C.F., and Hall, E.L., "A Nonlinear Model for the Spatial Characteristics of the Human Visual System", *IEEE Trans. Systems, Man and Cybernetics*, Vol. SMC – 7. No. 3, Mar. 1977, pp.161-170.

143. Fougeras, O.D., Digital Color Image Processing within the Framework of a Human visual Model, *IEEE Trans. ASSP*. Vol. 27, No. 4, Aug. 1979, pp.380-393.

144. Buchsbaum, G., An Analytical Derivation of Visual Nonlinearity, *IEEE Trans. Biomed. Eng.*, Vol. BME-27, No.5, May 1980, pp.237-242.

145. Roetling, P.G., Visual Performance and Image Coding, *SPIE/OSA*, Vol. 74, 1976, pp.195-199.

146. Dooley, R.P., Predicting brightness appearance at edges using linear and non-linear visual describing functions, pres. at *SPSE Annual Meeting*, May 14, 1975, Denver, Colorado.

147. Linsker, R., From Basic Network Principles to Neural Architecture, (series) *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 83, Oct-Nov 1986, pp.7508-7512; pp.8390-8394; pp.8779-8783.

148. Linsker, R., Development of feature analyzing cells and their columnar organisation in a layered self-adaptive network, in Cotterill, r. (ed) *Computer Simulation in Brain Science*, Cambridge Uni. Press, pp.416-431, 1988 (Conference in Copenhagen, Aug. 1986).

149. Linsker, R., Towards an organization principle for a layered perceptual network, in Anderson, d. (ed.), *Neural Information Processing Systems - Natural and Synthetic*, American Institute of Physics, N.Y. 1988.

150. Linsker, R., An application of the principle of maximum information preservation to linear systems, in *Advances in Neural Information Processing Systems 1*, C. Denver, Nov. - Dec. 1988, Touretzky, D.S., (ed.), Morgan-Kaufman, San Mateo, CA, 1989, pp.186-194.

151. Linsker, R., How to generate ordered maps by maximizing the mutual information between input and output signals, *Neural Computation*, 1, pp.402-411, 1989, MIT.

152. Linsker, R., *Designing a sensory processing system : what can be learned from principal components analysis*, IBM research report RC14983 (#66896) 9/27/89.

153. Miller,K.D., Keller,J.B. & Stryker,M.P., Ocular dominance column development: analysis and simulation, *Science*, Vol.245, 11 Aug. 1989, pp.605-615.

154. Marshall, J.A., "Self-organizing Neural Networks for Perception of Visual Motion", *Neural Networks*, Vol.3, pp.47-74, 1990.

155. Bossomaier, T., and Snyder, A.W., "Why spatial frequency processing in the visual cortex?", *Vision Res.*, Vol.26, No.8, 1986, pp.1307-1309.

156. Barlow, H.B., "The absolute efficiency of perceptual decisions", *Phil. Trans. R. Soc. Lond. B.*, Vol.290, 1980, pp.71-82.

157. Kersten, D., "Predictability and redundancy of natural images", *J. Opt. Soc. Am. A.*, Vol.4, No.12, Dec. 1987, pp.2395-2400.

158. Burkhardt,H. & Zisserman,A., *Proceedings of the ESPRIT BRA workshop on Invariants for Recognition*, held in S. Margherita Ligure, May 1992, in conjunction with ECCV2.

159. Hubel, and Wiesel, T.N., *J. Physiol, Lond.*, Vol.148, p.574 (1959); Vol. 160, p.106 (1962).

160. Field, D.J. and Tolhurst, D.J., The Structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex, *Proc. R. Soc. Lond. B*, Vol.228, pp.379-400 (1986).

161. Stone, J., Morphology and physiology of the geniculocortical synapse in the cat: The question of parallel input to the striate cortex, *Investigative Ophthalmology*, Vol.11, pp.338-46.

162. MacKay, D.M., Strife over visual cortical function (News and Views), *Nature*, Vol.289, 15 Jan. 1981, pp.117-118.

163. Campbell, F.W. and Robson, J.G., Application of Fourier analysis to the visibility of gratings, *J. Physiol*, Vol.197, pp.551-566.

164. Wilson, H.R. and Bergen, J.R., A four mechanism model for threshold spatial vision, *Vision Research*, Vol.19, pp.19-32.

165. Wilson, H.R. and Gelb, D.J., Modified line element theory for spatial-frequency and width processing, *J. Opt. Soc. Am. A*, Vol.1, No.1, Jan. 1984, pp.124-131.

166. Wilson, R. and Knutsson, H., Uncertainty and Inference in the visual system, *IEEE Trans. Systems Man and Cybernetics*, Vol.18, No.2, March/April 1988, pp.305-312.

167. Wilson, R. and Granlund, G.H., The Uncertainty Principle in Image Processing, *IEEE T. PAMI*, Vol.6, No.6, Nov. 1984 pp.758-767.

168. Wilson, R. and Spann, M., Finite prolate spheroidal sequences and their applications, Part I, *IEEE T.PAMI.*, Vol.9, No.6 pp.787-795, Nov. '87; Part II, *IEEE T.PAMI.*, Vol.10, No.2, March 1988, pp.193-203.

169. Wilson, R and Spann, M., *Image Segmentation and Uncertainty*, Wiley, U.K., 1988.

170. Burt, P.J., The Interdependence of Temporal and Spatial Information in Early Vision, in Arbib, M. and Hanson A., pp.263-277.

171. Van Doorn, A.J. and Koenderink, J.J., The Structure of the human motion detection system, *IEEE T.SMC.*, Vol.13, No.5, pp.916-922, Oct. 1983.

172. Logothetis, N.K., Schiller, P.H., Charles, E.R. and Hurlbert, A.C., Perceptual Deficits and the activity of the colour opponent and broad-band pathways at Isoluminance, *Science*, Vol.247, 12 Jan. 1990, pp.214-217.

173. Hammond, P. and MacKay, D.M., *J. Physiol. Lond.*, Vol.284, p.117, 1978.

174. Groos, G.A., Hammond, P. & MacKay, D.M., *J. Physiol Lond.* Vol.260, p.47, 1976.

175. Gibson,J.J., *The senses considered as Perceptual Systems*, Boston, Houghton Mifflin 1966, quoted in Marr [9, p.29].

176. Fodor,J.A. & Pylyshyn,Z.W., How direct is visual perception? Some reflections on Gibson's "Ecological Approach", *Cognition*, Vol.9, pp.139-196, 1981.

177. Ullman, S., Against Direct Perception, *Behavioral and Brain Sciences*, Vol.3, pp.373-415.

178. Marr,D. & Nishihara,H.K., Representation and recognition of the spatial organisation of three-dimensional shapes. *Proc. R. Soc. Lond.*, Series B, Vol.207, pp.187-217.

179. Marcelja, S., Mathematical description of the responses of simple cortical cells, *J. Opt. Soc. Am.*, No.11, Nov. 1980, pp.1297-1300.

180. Daugman, J.G., Two-dimensional spectral analysis of cortical receptive field profiles, *Vision research*, Vol.20, pp.847-856.

181. Woodward, P.M., *Probability and Information Theory, with Applications to Radar*, McGraw-Hill, N.Y. 1953.

182. Kulikowski, J.J. and Bishop. P.O., Fourier analysis and spatial representation in the visual cortex, *Experientia*, Vol.37, pp.160-163.

183. Sakitt, B. and Barlow, H.B., A model for the economical encoding of the visual image in cerebral cortex, *Biological Cybernetics*, Vol.43, pp.97-108, 1982.

184. Pollen, D.A. and Ronner, S.F., Visual Cortical Neurons as localized spatial frequency filters, *IEEE TRANS. SMC.*, Vol.13, No.5, Sept/Oct. 1983, pp.907-916.

185. Webster, M.A. and De Valois, R.L., Relationship between spatial -frequency and orientation tuning of striate-cortex cells, *J. Opt. Soc. Am A*, Vol.2, No. 7, July 1985, pp.1124-1132.

186. Watson, A.B., Efficiency of a model human image code, *J. Opt. Soc. Am. A*, Vol.4, No.12, Dec. 1987, pp.2401-2417.

187. De Valois, R.L. Albrecht, D.G. and Thorell, L.G. Orientation selectivity of cells in cat striate cortex, *J. Neurophysiol*, Vol.37, pp.1394-1409, 1974.

188. Morshon, J.A., Two-dimensional spatial frequency tuning of cat striate cortical neurons, *Soc. Neurosci Abstr.* 9th Annual meeting, Atlanta, Ga., 1979, p799.

189. Hawken, M.J. and Parker, A.J., Spatial properties of neurons in the monkey striate cortex, *Proc. R. Soc. Lond. B*, Vol.231, p.251-288, 1987.

190. Daugman, J.G. and Kammen, D.M. Image Statistics, Gases, and Visual Neural Primitives, IV - 163 - IV - 175.

191. Marr, D. and Nishihara, H.K., Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. R. Soc. Lond. B*, Vol.200, pp.269-294.

192. Barlow,H.B., Sensory mechanisms, the reduction of redundancy, and intelligence, in *Proceedings of the National Physical Laboratory Symposium on the Mechanization of Thought Processes*, H.M. Stationary Office, London, 1959, No.10, pp.535-539.

193. Barlow,H.B., Perception: what quantitative laws govern the acquisition of knowledge from the senses?, in *Functions of the Brain*, Wolfson College Lectures, C.Cohen (ed.), Clarendon, Oxford, 1985, pp.11-43.

194. Duda, R.O. & P.E. Hart, *Pattern Recognition and Scene Analysis*, Wiley, NY, 1973.

195. Ballard, D.H. & Brown, C.M., *Computer Vision*, Prentice-Hall, 1982.

196. Attneave, F., Informational aspects of visual perception, *Psych. Rev.*, Vol.61, pp.183-193, 1954.

197. Blakemore, C., *The Mind Machine*, BBC Books, U.K., 1988.

198. Bickerton, D. quoted in [197], p.181.

199. Quoted in [19, p.510]; originally from Peirce, C.S. *Collected Papers*, Vol 3, Harvard University Press, Cambridge, Mass, 1960.

200. Bezdek,J.C. & Pal,S.K., *Fuzzy models for Pattern Recognition: methods that search for structure in data*, (Selected reprint volume of IEEE Neural Networks Council), IEEE Press, NY, 1992.

201. Wilson, R., and Knutsson, H., Uncertainty and Inference in the visual system, *IEEE Trans. SMC*, Vol.18, March/April 1988, pp.305-312.

202. Wilson, R., and Granlund, G.H., The Uncertainty Principle in Image Processing, *IEEE Trans. PAMI*, Vol. ?? No.6, Nov. 1984, pp.758-767.

203. Grossberg,S., *Neural Networks and Natural Intelligence*, MIT Press/Bradford, 1988.

204. Attneave,F., Informational aspects of visual perception, *Psych.Rev.*, Vol.61, pp.183-193, 1954.

205. Holland,J.H., *Adaptation in natural and artificial systems*, Uni. Michigan Press, Ann Arbor, 1975.

206. Goldberg,D.E., *Genetic algorithms is search, optimization and machine learning*, Addison-Wesley, 1989.

207. Holland,J.H., Genetic algorithms and classifier systems: foundations and future directions. *Proc. ICGA2*.

208. Bersini,H. & Varela,F., Hints for adaptive problem solving gleaned from immune networks, in Schwefel,H.-P. & Manner,R., *Parallel problem solving in nature*, Lecture notes in computer science No. 496, Springer-Verlag, Berlin, pp.343-354, 1991.

209. Aubin,J.P., *Viability theory*, Birhauser, Berlin, 1991.

210. Anderson, A.. Learning from a computer cat, *Nature*, Vol.331, 25 Feb 1988, pp.657-659 (News & Views).

211. Zipser, D. and Andersen, R.A., A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, Vol.331, 25 Feb. 1988, pp.679-684.

212. Lehky, S.R. and Sejnowshi, T.J., Network model of shape-from-shading neural function arises from both receptive and projective fields, *Nature*, Vol.333, 2 June 1988, pp.452-454.

213. Grossberg,S. & Mingolla,E., A neural network architecture for pre-attentive vision, *IEEE T. Biomedical Engineering*, Vol.36, No.1, 1989, pp.65-84.

214. Burr,D.C., Ross,J. & Morrone,M.C., Seeing objects in motion, *Proc R. Soc. Lond.*, Series B, Vol.227, pp.249-256, 1986.

215. Purves,D., The trophic theory of neural connections, *Trends in Neuroscience,* October 1986, pp.486-489.

216. Landmesser,L., Axonal guidance and the formation of neural circuits, *Trends in Neuroscience,* Oct. 1986, pp.489-492.

217. Oppenheim,R.W., The neurotrophic theory and naturally occurring motoneuron death, *Trends in Neuroscience,* Vol.12, No.7, 1989, pp.252-255.

218. Mastronarde, D.N., Correlated firing of retinal ganglion cells, Trends In Neuroscience, Vol.12, No.2, 1989, pp.75-80.

219. Stryker, M.P., Is grandmother an oscillation? (News & Views), *Nature,* Vol.338, 23 March 1989.

220. Bressler, S.L., The Gamma wave; a cortical information carrier?, TINS, Vol.13, No.5, 1990, pp.161-162.

221. Logothetis, N.K. and Schall, J.D., Neuronal Correlates of Subjective Visual Perception, *Science,* Vol. 245, 18 Aug. 1989, pp761-763.

222. Sirovich, L. and Kirby, M. (1987). Low dimensional procedure for the characterization of human faces, *J.Opt.Soc.Am.* Series A, Vol.4, No.3 (March), pp.519-524.

223. Kirby, M. and Sirovich, L. (1990). Application of the Karhunen-Loève Procedure for the Characterization of Human Faces, *IEEE T.PAMI.*, Vol.12, No.1, (January).

224. Lippmann, R.P. (1987). An Introduction to Computing with Neural Nets, *IEEE ASSP Magazine*, April, pp.4-23.

225. Linsker, R. (1988). Development of feature-analyzing cells and their columnar organization in a layered self-adaptive network, pp. 416-431, in *Computer Simulation in Brain Science,* Cotterill, R.M.J., Cambridge University Press (from conference on Copenhagen, Aug., 1986).

226. Lehky, S.R., and Sejnowski, T.J. (1988). Network model of shape-from-shading: neural function arises from both receptive and projective fields, *Nature,* Vol.333, 2 June 1988, pp. 452-454.

227. Linsker, R. (1989). *Designing a Sensory Processing System: What can be learned from Principal Components Analysis?,* IBM Research Division, Research Report RC14983 (#66896), 9/27/89.

228. Oja, E. (1982). A Simplified Neuron Model as a Principal Component Analyzer, *J. Math. Biology,* Vol.15, pp. 267-273.

229. Jain, A.K. (1989). *Fundamentals of Digital Image Processing,* Prentice-Hall.

230. Duda, R.O. and Hart, P.E. (1973). *Pattern Classification and Scene Analysis,* Wiley, NY.

231. Therrien, C.W. (1989). *Decision, Estimation and Classification*, Wiley.

232. Golub, G.H. and Van Loan, C.F. (1989). *Matrix Computations*, 2nd Ed., The Johns Hopkins University Press, Baltimore, Md.

233. *Matlab User's Guide*, (1989). The Mathworks, Inc., 21 Eliot St., South Natick, MA 01760.

234. Dongarra, J.J., Moler, C.B., Bunch, J.R., Stewart, G.W. (1979). *LINPACK User's Guide*, SIAM, Philadelphia.

235. Gunning, J., "Automatic Feature Selection for Automated Visual Inspection", *M.Eng. Thesis*, Dublin City University, 1991, (supervised by N. Murphy).

236. Dayhoff, J.E., 1990, *Neural Network Architectures - An Introduction*, Van Nostrand Reinhold, New York.

237. Wasserman, P.D., 1989, *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York.

238. Karnin, E.D., 1990, *IEEE Trans. Neural Networks*, Vol.1, 239-242.

239. Touretzky, D.S. and Pomerleau, D.A., 1989, *Byte*, Aug., 227-233.

240. Schalkoff, R., 1992, *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley.

241. Whitaker,R. *The Observer* (An electronic forum for those interested in Autopoiesis, Enactive Cognitive Science and related matters), No.2, 1993.