Phil Trans Roy Soc B 352, 1117–1120 (1997)

INTRODUCTION

Geoff Sullivan had the idea for a meeting on 'knowledge-based vision' many years ago, and was absolutely delighted when the Royal Society gave the go-ahead for detailed planning to start. He was the initiator and guiding spirit, and his original co-organizers, Richard Gregory and Horace Barlow, spent most of their time simply agreeing with his suggestions. Then in the summer of 1996 he told us the news that he had cancer of the lung, and a few weeks later, that it was inoperable. He kept going for another month, buoyed up by the hope that radiotherapy would slow its progress, but it was ineffective and he died at the end of August, showing enormous courage and working until the very end. James Anderson, a long-term colleague in the Computer Science department, took over the reins when Geoff died, and he has worked wonders in maintaining the momentum that Geoff originally imparted, while adding a spin of his own.

The aim of the discussion meeting was to promote an exchange of views between three research communities—those concerned with the physiological mechanisms of sensation and perception, those concerned with their psychology, and those who devise computational methods for achieving what we do with such deceptive ease when we look at the world around us.

The focus of the meeting was to explore how sensory data and perceptual knowledge can be brought together to achieve vision, but in the planning stages we had a heated argument about the nature of 'knowledge' and where it comes from. One view, favoured by psychologists and some of the computer-vision community, was that knowledge is imported from outside and is applied in a 'top-down' fashion to make sensory signals intelligible, since these are hopelessly incomplete by themselves. The other is favoured by those who are becoming acutely aware that the visual system is a profusely interconnected, multi-layered, heterarchy that does not have a 'top', and also by the growing section of the computer-vision community that appreciates what a rich mine of valuable knowledge is contained within the statistical structure of sensory messages themselves. According to this view, knowledge is partly inherent in the innately determined anatomy of the visual system, but is supplemented to an important extent by analysis of the associative structure of the messages being handled.

One can regard the two main streams of thought about perception in psychology as championing these two rival sources of knowledge. More than 100 years ago, von Helmholtz (1821–1894), introduced the notion that perception depends heavily on unconscious inference (see von Helmholtz 1925). Although he retreated from this position under the onslaught of philosophers who argued that inference was necessarily conscious, he continued to maintain that there was an active, intelligent, problem-solving, aspect in the interpretation of the necessarily incomplete messages about the world that are provided by the senses. In contrast Gibson (1950) drew attention to the directness of perception and the way it is instantly triggered by the whole pattern of the scene before the eyes. Even though his ideas are more recent it is not altogether easy to see what he is getting at, but perhaps he is emphasizing that much of our perceptual apparatus is innately determined and exploits universal features of visual patterns such as optic flow.

Over the past 20 years it has become clear that computational vision can add a great deal to these two classical psychological viewpoints. Neither introspection, nor psychological experiment, nor physiological analysis provide the same insights as those obtained by actually trying to do what the eye does for us so effortlessly, and the natural difficulties encountered when attempting these tasks have been slowly coming to light. The computational papers in this volume present a cross-section of the struggles being made by many different groups to imitate various aspects of vision, and they record the progress being made. It is rather as if hordes of climbers were all tackling the same mountain; one group finds itself stuck in a gulley, another has accidentally climbed the wrong peak, and a third is rapidly approaching a quite unclimbable rock face. One would need a wide historical knowledge to know why each group finds itself where it is, but from these many efforts a rather detailed picture is emerging of what is involved in performing the task of vision, and in particular these efforts are showing what parts are difficult.

There is no single problem; biological visual systems just seem to be devilishly good at obtaining useful knowledge of the world from the light that strikes the eye, and it is a genuinely hard task to emulate them. One of the general points to emerge is that there are statistical limits to what the eye can do, set by random and uncontrolled variations in the signals generated by objects and events in the world around us, and the important point is that, once these limits are understood, the phrase 'devilish good' can in principle be expressed quantitatively. The statistical nature of many problems in vision was not evident to early psychologists and physiologists, and it is perhaps the single most important point to emerge from the computational approach.

The view of the problem from statistics and information theory actually throws much light on the relation between intelligence and knowledge in vision. Intelligence is the active, problem-solving aspect that von Helmholtz (1925) drew attention to, but it now emerges that knowledge is not only provided by the Gibsonian, innately determined, structural mechanisms of the perceptual apparatus (Gibson 1950), but is also added to by the active process. This follows from two insights. The first is the close relation between

intelligence and the identification of structure or redundancy in a set of messages (Barlow 1983). This is shown by considering what is required to do well in an intelligence test: for one type of problem one must identify a pattern or structure in a series of items, and the solution is obtained by continuing that pattern; the solution to other problems results from having applied this paradigm in one's everyday experience. The second insight is that structure or redundancy (in the information theory sense) constitutes knowledge (Barlow 1996). This follows directly from the fact that both of them define some way in which a set of messages deviates from complete randomness.

These insights from information theory tell one that intelligence is the process of identifying redundancy, and that knowledge is the store of identified redundancy that this process builds up. This resolves the paradox, pointed to by Gregory (1987), that one thinks of knowledgeable people as being intelligent, yet the more knowledge one has, the less intelligence is likely to be needed for solving a problem; he names these 'kinetic intelligence' and 'potential intelligence' by analogy with creating and storing energy, though information is actually related to entropy, not energy. These issues are highly relevant to the considerable fraction of papers in this volume that use statistical methods to supply some of the knowledge used in vision.

The papers are published in the order in which they were given at the meeting. Richard Gregory first presents his viewpoint as a psychologist in 'Knowledge in perception and illusion', and he introduces a new classification of illusions, which have provided so much insight into the psychology of perception. James Anderson gives the second paper, on 'Representing geometrical knowledge', which explains a mathematical technique for handling visual images that is not only useful in computer vision, but also represents very naturally and economically the ways in which we can manipulate visual images in our minds. Horace Barlow ends this introductory trio with a paper on 'The knowledge used in vision and where it comes from', which argues that adaptive mechanisms discount established properties of the input signals, including associative properties. This use of the redundancy in sensory messages can thereby improve the ability of the system to detect new associations that are likely to have significance for the survival of an animal.

The paper by Matteo Carandini and his colleagues on 'Adaptation to contingencies in macaque primary visual cortex' shows that the adaptation of single cortical neurones does not always depend simply on their degree of activation, but can be selective for a contingency in the input that does not correspond to the neurone's selectivity of response. This supports the hypothesis that knowledge is extracted from the redundancy of sensory messages. The paper by Dan Kersten on 'Perceptual categories for spatial layout' gives many examples in which shadows are used to assign distance to moving objects in movie sequences. Could the link between an object and its shadow be established by an associative process?

The theme that the redundancy of sensory messages contains valuable knowledge was continued later, but the next paper, by David Cliff and Jason Noble on 'Knowledge-based vision and simple visual machines', brought technology and neurology into conflict with theory. In both fields the notion of internal representations is central, but Cliff questions whether it is useful to think of neural representations at all. He points out that many simple visual systems seem to perform their job without it being possible to identify, in the neuronal machinery, any correlates of the abstractions that computer scientists and psychologists postulate.

The paper by Geoff Hinton and Zoubin Ghahramani on 'Generative models for discovering sparse distributed representations' discusses learning in a new type of artificial neural network that develops interconnections within a layer of artificial neurones to explain away redundant correlations in the input data. The strengths of these interconnections represent an increase in knowledge in Barlow's terms, and they neatly explain the occurrence of topographical maps of neurones in the visual areas of the brain. This throws a challenge back to Cliff and Noble: if their results with simple 1D stereo vision generalize to real images, then we can expect Hinton's neural nets to develop a topographic map of the image, providing exactly the sort of representation of knowledge that Cliff says he cannot find in neural networks.

The paper by Shimon Edelman and Sharon Duvdevani-Bar, 'A model of visual recognition and categorization', discusses the relationship between the recognition of objects that have been seen before and classification of new objects. In essence, all objects are classified by their distance from known prototypes. If the distance is small then the new shape is usually put into the same category as the other prototypes, but if the position of the new shape straddles existing category boundaries or is very distant from any prototype, then the new object is simply categorized as being more or less like the nearest few prototypes. The authors argue that in many everyday circumstances recognition of object classes is more useful than recognition of individual objects, so that classification should at least be computed before individual identity.

In their paper, 'Neurocomputational bases of object and face recognition', Irving Beiderman and Peter Kalocsai compare the recognition of faces with recognizing other objects. Special strategies have to be used for face recognition because faces are very similar to each other, yet identification of individuals is particularly important. Perhaps surprisingly, these break down in unusual conditions, such as the reversed contrast of photographic negatives.

The paper by Chris Frith and Raymond Dolan on 'Brain mechanisms associated with top-down processes in perception' describes recent studies with modern brain imaging systems, which can form images of the brain with a spatial resolution of 5–15 mm and a temporal resolution of 10–30 s. This raises the exciting possibility of seeing how the parts of the brain that are active vary with the task being performed, though this enterprise is beset with difficulties. In one elegant experiment people were first shown a degraded version of a photograph, then shown the original photograph, and finally the degraded photograph again. On the first presentation of the degraded photograph, people were unable to recognize what was in it, but on the second presentation people were able to recognize the object by remembering the original, non-degraded photograph. Comparing the brain images for the two occasions on which people looked at the degraded images highlights the brain areas that are active when supplying knowledge of the remembered object.

Mike Land and Sophie Furneaux's contribution, 'The knowledge base of the oculomotor system' describes experiments in which a new eye-tracking system enabled eye movements to be recorded while the subjects were performing various realistic tasks, such as driving a car, playing table tennis, or reading music. They produce evidence that the eyes search actively for needed information and store what they find in short-term storage buffers. Their results suggest how perceptual hypotheses fill gaps in sensed data and enable prediction into the immediate future, so that motor behaviour is produced at the appropriate moment in spite of the physiological delay-times of sensory signals and effectors.

The robotics expert Mike Brady writes on 'The forms of knowledge mobilized in some machine vision systems'. He looks at knowledge-based machine vision for such practical uses as medical diagnosis in mammography, where image improvement is achieved by providing (or allowing the machine to learn) characteristics of what is being recorded.

David Milner's paper on 'Vision without knowledge' develops the radically new idea that visual processing for controlling motor behaviour may be different from processing for perception—with different knowledge bases. There is a gross disparity between the errors for a given task when these are assessed by a subject's behavioural skill and by his or her conscious knowledge. Clinical findings from a patient with visual-form agnosia are also described: she cannot recognize objects even though her visual acuity is normal.

Aaron Bobick's paper on 'Movement, activity, and action: the role of knowledge in the perception of motion' describes a computational approach analysing and identifying the movements and actions in a video sequence. He uses a computer program that transforms motions into a single image where the most recently moving pixels are brightest. This provides an explicit temporal pattern of movement from which the direction of motion can be deduced, and actions can often be identified in spite of the enormous reduction in the amount of information presented. The program also learns implicit statistical relationships between these movements in terms of hidden Markov models. The system is used to recognize gestures in American sign language.

The paper by Chris Taylor and colleagues on 'Model-based interpretation of complex and variable images' argues that vision is an essentially statistical problem and shows how the range of normal variability of objects in images can be allowed for and used to aid interpretation. The range of legal variation in objects is captured by principal component analysis (PCA) in the appropriate image space, and with other forms of statistical analysis it is possible to separate out the different sources of variation, such as the changes in facial appearance according to the individual, the lighting, the direction of gaze and facial expression. The practical uses of these techniques in the interpretation of medical images is illustrated.

Different kinds of knowledge are also discussed in the paper 'Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy' by Glyn Humphreys, Jane Riddoch, and Cathy Price. They use brain imaging and psychological experiments motivated by neurophysiological theories to tease apart the various kinds of knowledge that go together to name objects and put them into meaningful categories. They find that knowlege is not processed serially, but that knowledge cascades from early visual processing to late semantic interpretation, with processing starting in some stages before it is complete in others. They also report evidence of semantic priming effects. This reinforces the view that the brain supports a truly complex heterarchy of visual knowledge and processing.

Alex Pentland's paper on 'Content-based indexing of images and video' describes a remarkably powerful system for recognizing and categorizing images, including moving sequences. The system recognizes objects by using the most significant terms of an encoding that could reconstruct the whole image. The system currently uses PCA and is guided toward an efficient recognition of individual human faces by finding landmarks, such as the eyes, mouth, and nose. The image is then warped so that the landmarks appear in a standard position. This gives the PCA, which is a statistical process, the best chance of finding a correct match. Thus a computer scientist would say that high-level knowledge of landmarks on a human face is provided by the programmer and is used to guide a bottom-up process. The papers by Horace Barlow and Richard Gregory in this volume debate the nature of this kind of knowledge in their respective treatments of Gregory's hollow head illusion.

It is notable that the papers by Edelman, Bobick, Taylor and Pentland all depend strongly on PCA (or Karhunen-Loeve Analysis), showing that the computational vision community appreciates that knowledge

can be derived from the associative structure of sensory messages. But in general, statistical structure of a higher order than pairwise associations has not been used; only Hinton's work explicitly makes use of higher order associations.

The usefulness of statistical structure in performing visual tasks was recognized by the authors of many of the papers, but there was a strong tendency for others to assume that knowledge means conscious, cognitive knowledge, and that it is applied from the 'top'. The heterarchical organ that actually does the seeing does not, however, have a recognizable 'top', and much knowledge is either embodied in the structure of the visual system, or available from the analysis of redundancy in neural signals themselves. The phrase 'top-down' covers many types of operation and should be re-examined whenever it is used.

Originally we intended there to be a logical progression in the papers, starting with those having a strong emphasis on knowledge derived from the input, and ending with problems where the cognitive element was greater, but this organization and sequence may appear a bit confused because we deliberately mixed the biological and physical viewpoints, and the content of the papers often did not correspond precisely with our expectations. The lively discussions at the meeting suggested that considerable cross-cultural comprehension had been achieved, and as editors, we have tried to reinforce this by asking that the main points of the papers be made intelligible to both biologists and physicists. We hope some of the cross-cultural comprehension evident at the meeting can be gleaned from this written record.

April 1997

James Anderson
Horace Barlow
Richard Gregory

REFERENCES

Barlow, H. B. 1983 Intelligence, guesswork, language. *Nature* **304**, 207–209.

Barlow, H. B. 1996 Banishing the homunculus. In *Perception as Bayesian inference* (ed. D. Knill & W. Richards). Cambridge University Press.

Gibson, J. J. 1950 *The perception of the visual world*. Westport, CT: Greenwood Press.

Gregory, R. L. 1987 Intelligence based on knowledge–knowledge based on intelligence. In *Creative intelligences* (ed. R. L. Gregory & P. K. Marstrand). London: Francis Pinter (for The British Association for the Advancement of Science).

Helmholtz, H. von 1925 *Physiological optics. III. The theory of the perceptions of vision* (translated from 3rd German Edition, 1910). Washington: Optical Society of America.