

## WILL SEEING MACHINES HAVE ILLUSIONS?

---

R. L. GREGORY  
PSYCHOLOGICAL LABORATORY  
UNIVERSITY OF CAMBRIDGE

The ability of the higher animals to accept and interpret information from distant objects confers enormous advantages for creatures (or machines) which respond only to immediate stimulation and have no opportunity to anticipate the future. Distance receptors, especially the eyes, serve as early warning systems by giving information of distance events, making it possible to gauge the probable future. The classical biological notion of stimulus-response applies to creatures limited to touch information. The development of distance-receptors evidently allowed brains to develop to give strategic behaviour. It is unfortunate that the early, now classical, studies of reflexes involving touch and the internal regulation of the body have been so largely taken over to describe brain function, for these concepts are inadequate for describing the central nervous system. They tell nothing about how brains handle information from the eyes, to allow animals and man to see. They tell us nothing about decision-making: how present experience is related to the stored past to predict the immediate future. Reflexes imply an essentially passive system—a system which would be quiescent in the absence of sensory inputs—but this is far from the case with the higher living systems.

Biological information processing is almost incredibly efficient. The general slowness of animal movements belies the rate of perceptual decision-taking, which is even more remarkable when we consider that sensory information is transmitted with pulse rates always under a kilocycle and generally but a few hundred cycles maximum. An important reason for the speed and efficiency of the visual system is its very large number of parallel channels—at least  $10^6$  for the human eye. Sequential processing is inevitably slow (which handicaps the ear) but the eye, though limited by the limited band-width of its channels, is faster than any man-made computer, though these have electronic components several orders more rapid in operation.

The brain interprets two-dimensional images in the eyes as representing objects distant in space. The more one considers what is involved, the more astonished one becomes at the ability of nervous systems to recognise objects and judge distance and speed in the ever-changing external world. Simultaneous parallel transmission of information is however but the beginning of the story: just as important are various 'strategy' short cuts in processing. These short cuts are not always successful, and can produce serious errors. Some visual errors are of little interest, but others are worth studying in detail for they are clues to the way the visual system functions. Some errors are due to *limitations of the components* of the nervous system—for example fatigue or adaptation of cells with prolonged use. Others are due to the *logical difficulties of the problems* to be solved in seeing the world. These logical difficulties are of general interest, for we should expect to encounter them in *any* system, including man-made seeing machines, set to solve the problem.

#### THE PROBLEMS FOR SEEING MACHINES

Biological seeing machines handle spatial patterns which must be classified, stored and compared. Some kind of filing system with ready access is required, and in human perception the perceptual 'filing cards' contain a wealth of information going far beyond purely sensory characteristics. To say, 'I see a table', is to go beyond, 'I see a brown rectangular shape, with square vertical protrusions extended below it . . .'. To see a table is to know that it has an underneath; that the 'square protrusions'—the legs—support it, and are necessary for its stability. It is to know that it will not bark, or explode, that it may be used for resting things on, that it is easily scratched, heavy to lift and that one may hurt one's knee on hidden edges. When we see a table, we 'see' a host of non-visual things—past and future, as well as present. We 'see' that it was man-made, and that it will not 'softly and suddenly vanish away . . .'.

The point is that to *see* a table is far more than to have a *picture* of a table. A picture is but a part copy of the original, and is nothing until interpreted.

A seeing machine is far more than a photographic or television camera. It translates optical information into appropriate action (be this limited to linguistic description) and this involves assessing the significance of the pictures in the eye.

Looking at a *picture* of a table is a most curious business. The visual input may even be identical to the original—truly excellent stereoscopic coloured cinema—and yet the situation is not equivalent to that of seeing a real table. Though the visual inputs are the same, the information on the internal 'file cards' for 'object' and for 'picture of object' are quite different. We cannot *use* this picture table. We cannot put things on it, or hurt our knee by banging it. Even a perfect picture is perceptually very different from the object itself, because a picture has quite different, if any, causal properties. And yet the retinal image is only a picture.

## GREGORY

To identify an object a perceptual system equipped with filing cards has merely to recognise the *minimum* defining characteristics. It may use probabilities; it may 'consider' first the most likely alternatives, so that there is no need to hunt through the whole, or a large part, of the total file. But this approach to saving computation time must lead sometimes to misrecognition of improbable objects. It may also give spurious addition of characteristics which 'ought' to be present. It is said that people shown aircraft for the first time 'see' feathers on their wings. Feathers 'ought' to be present, and they are added from the data on the file card labelled 'wing', if all previous wings had feathers.

In studying the rare case of a man who recovered sight (by corneal grafting) after being blind since infancy, I noticed many times that he would correctly identify objects previously known to him by touch from quite small distinguishing features. He did not seem to see things completely, but generally made do with very little information provided he was familiar with the object from previous touch experience. When it came to drawing, he was incapable for a year or more of representing features not previously familiar to him. For example, he could not draw the front of a bus—and this he had never touched as too dangerous—though he could draw the rest quite well (Fig. 1), Gregory and Wallace (1963).

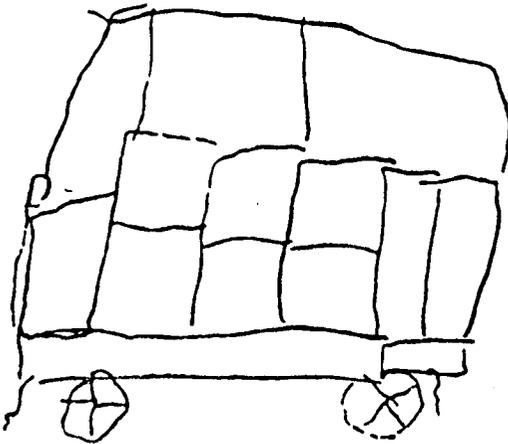


FIG. 1. Drawing of a bus (48 days after the first operation).

The ability of the human perceptual system to extrapolate way beyond the immediate sensory evidence is clearly demonstrated by perception of *imperfect* pictures. Consider cartoons: a few lines . . . we 'see' a complete face, expression and all (Fig. 2). But the blind man who recovered his sight made nothing of cartoons. He lacked the appropriate internal file cards.

Recently, electrophysiological studies of visual processes in lower animals have shown that but few characteristics are coded from the retina and presented to the central nervous system. In the case of the frog, Lettvin,

Maturana, McCulloch & Pitts (1959) found that movement, angles, corners and intensity of light are almost the only features of the world which are signalled to the frog's brain. It is a simple language, but for the frog generally adequate. It has been suggested by W. H. Rushton (1953) that the fovea of the human eye presents information of all pattern, brightness, colour and movement almost completely to our brain, while the peripheral regions of the retina are more selective, as in primitive eyes. As Gordon Walls (1963) points out, the more primitive the brain the more complex the eye. Primitive

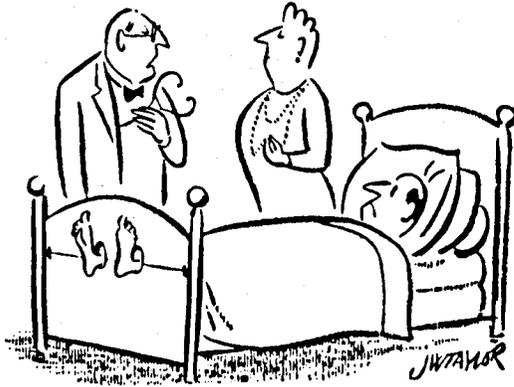


FIG. 2.  
 '... and keep him in bed for another day or two.'  
 Drawing reproduced by permission of *Punch*.

eyes are complex in order to transmit only what is necessary for the limited brain computer. Primitive brains have few, and built in, file cards while the higher mammals have simple eyes, but vastly complex brains which develop cross-indexed file cards during the long period of helpless infancy and childhood. It now seems clear that for 'perceptual learning' to occur, there must be repeated active movement with associated visual signals, which come to represent external objects through lengthy experiments with real objects. Vision does not stand by itself; it needs the more direct information from active touch for the signals from the eyes to have meaning, to give perception of the external world of objects. Retinal images are not stimuli producing reflex action—they are more like references to the perceptual file cards, built up by active experiment and conflict with objects.

### SEEING OBJECTS IN SPACE

Stereoscopic vision is rare in Nature, and is never of use at distances beyond a hundred feet or so. It is used for precise location of near objects by arboreal and some predator animals. Of far more general importance is the remarkable ability of all large brains to interpret the two-dimensional images of its individual eyes in terms of external three-dimensional space. The problem set is acute, because two-dimensional projections of three-dimensional space are infinitely ambiguous. Retinal images are subject to

an indefinite number of possible interpretations. Consider the simple case of an elliptical-shaped image. It could represent a tilted circle, an ellipse of a certain eccentricity located normal to the line of sight, or any of an infinite set of ellipses of given eccentricity and corresponding angle of tilt. Further, it may be a large distant object, or smaller but correspondingly nearer, to give the same image at the eye. An infinite set of objects lying at various angles and distances give the same image. The brain has to 'decide' which possibility to adopt for perception. When it decides wrongly, we suffer an illusion.

There is generally a wealth of available information—though none entirely reliable—for settling the size and distance of external objects, with sufficient precision for normal use. As is well known, the visual system makes use of a host of 'depth cues', such as gradual loss of detailed texture with increasing distance, haziness due to the atmosphere and nearer objects partly hiding those more distant. These cues were discussed in the nineteenth century by the great von Helmholtz (1925), who fully realised their importance, and they have been the subject of many investigations since, especially by J. J. Gibson (1950). Whatever the richness of depth cues, however, the visual input is always ambiguous. Though the brain makes the best bet on the evidence—it may always be wrong.

The kind of mistakes which occur when the bet is on the favourite though the favourite is not placed, is shown most dramatically by the demonstrations of Adelbert Ames (1946). The most impressive demonstration is given simply with a room which is non-rectangular, but so shaped that it gives the *same retinal image* as a rectangular room to an eye placed in a certain position. Now clearly this room, though queer shaped, *must* appear the same as a normal rectangular room, for it gives the same image to the eye. But consider what happens when objects are placed inside the Ames room. The further wall recedes at one side, so that an object or person standing in one corner is actually at a different distance than is a second object placed at the other far corner. These objects (or people) appear, however, to be at the same distance—and they are seen the *wrong size*. This is clear evidence that we assume rooms to be rectangular (because they usually are) and we interpret the size of objects according to their distance as given by this assumption. When the assumption is wrong we see wrongly. What Ames did was to rig the odds, and then we make the wrong decision on size and distance. A child may appear larger than a man. We may know this is absurd and yet continue to see a bizarre world. The retinal image is all right, but the odds have produced the wrong internal file cards and then the human seeing machine is upset, and gives a wrong answer.

It is interesting that the Ames room is seen correctly by peoples, such as the Zulus, brought up in a 'circular culture' of beehive huts where there are few reliable perspective features, such as rectangular corners and parallel lines, in their visual environment. To the Zulus, the odds are not rigged by the Ames room—to them this is not misleading perspective. They are not

subject to this illusion, but accept the room as the shape it is, and see the objects in it correctly in distance and size. This is a matter of very real importance. It shows that when we are transferred to an alien or bizarre environment, where our filing cards are inappropriate, we interpret the images in the eyes according to principles found reliable in the previous, familiar world—but now they may systematically mislead and then perception goes wrong. Space travellers beware!

The human perceptual system largely compensates for the geometrical shrinking of retinal images with object distance, to give approximate size-distance invariance, known as size constancy. It has recently been suggested, by the author, that size constancy should be attributed to two quite different active scaling processes. One scaling process functions simply according to seen distance. The other is set, or triggered, directly by depth cues—especially perspective features. It seems that we may attribute the well-known 'visual illusion' distortions to the direct setting of this constancy scaling mechanism by perspective features in the illusion figures. In every case, perspective features indicating *distance* produce illusory *expansion*. Now this perceptual expansion with indicated distance would be appropriate if the perspective angles were in fact due to different distances—as in retinal images of the truly 3-D world—but the illusion figures *lie on flat planes*, and so the size scaling induced by their perspective is inappropriate. What is appropriate for the retinal images of the 3-D world is inappropriate for the identical retinal images of 2-D projections of the world. Constancy scaling is inappropriate for pictures, for they are in fact flat, though they have depth cues (Gregory 1963, 1966).

The relation between touch and vision is important, but it has received little investigation until recently so that rather little is known. On general grounds, one can say that vision must ultimately depend on active exploration of the world by immediate contact, since retinal images must somehow come to represent the real world though they are but indirectly related and indeed infinitely ambiguous. This is borne out by the work of R. Held & A. Hein (1963). Current evidence suggests that in man vision is greatly dependent on infant experience, though logically it is possible that retinal patterns may have immediate significance through genetic coding derived from ancestral correlations between vision and touch. It is however an open question how much a man-made seeing-machine should be allowed to learn for itself. If it is to see unusual, or unique, objects it must be capable of some individual learning. We find that animals which are highly adaptable to a variety of environments require a long period, apparently for perceptual learning, before they can respond, except in a stereotyped and reflex manner, to external objects. We should expect the same for man-made seeing machines.

Consider the simple stimulus-response system; or a machine in which the output is directly determined by the state of its inputs. Such systems have a basic limitation—they cannot function with zero or with strictly inadequate

input. To take an example: a car driver can very often continue steering without disaster for several seconds if he is blinded by oncoming lights, or if he sneezes. He steers according to his prevailing 'hypothesis' of the position of the road ahead. He will also base his decisions on what other drivers are likely to do in the immediate future—not merely on what they are doing, as signalled by his eyes. This ability of animals to keep going in the temporary absence of sensory information is vital to their survival, and would seem an essential characteristic of an adequate man-made seeing machine. It involves the notion that seeing is not bound tightly to the current state of affairs as signalled to brain or computer. We may think of perception of the world as ever-changing running hypotheses, suggested and tested by sensory data but never determined wholly by current information from the senses. When sensory information is particularly inadequate, the 'perceptual hypothesis' will diverge from the truth, but generally it will be better than nothing in familiar environments. In strange environments we must expect the normally useful stock of hypotheses to be inappropriate and misleading. Indeed, a refresher course of perceptual learning will be required to cope with it. We know, from studies with distorting spectacles, that the adult human perceptual system is capable of radical readjustment to transformed conditions. This is generally useful, but it follows that perceptual systems are not good at absolute judgements, for we become adapted to changing conditions and lose our ability to tell that they have changed.

The nervous system and especially the sensory systems, apart from those of pain, are generally a.c. coupled. This has obvious advantages: it makes circuits more reliable in the face of internal noise and drift; it emphasises signals indicating change, which are of biological importance, but it can give rise to certain errors. For example the well-known 'waterfall effect'. Prolonged viewing of systematic movement leads to reduction in apparent velocity, and to a curious after-effect when the movement ceases, when illusory movement in the opposite direction is seen. Now it happens that this illusory movement is logically paradoxical—we see movement and yet we see no change in position. This is quite a common feature of sensory illusions, and may be attributed to a state of affairs where one out of two or more parallel channels is disturbed. The false signal may be accepted together with the other, correct, signals—then we experience a logical paradox—a physical impossibility. In the case of the paradoxical 'waterfall' illusion, it seems that movement can be signalled both (i) by change of position in a given time, and also (ii) by direct velocity signals from the retina. (This is analogous to velocity indicated by instruments such as a speedometer which make no reference to a clock.) If either system is upset we may experience a paradox—velocity without changes of position in time. We should expect any machine having parallel channels to produce paradoxes of this kind, when some channels are upset, or receive misleading information. This is fairly common when different senses disagree, but generally one

sense dominates over others so that errors are simply discarded. In man, vision dominates over auditory direction-finding, so that sound generally appears to come from where the source is seen. Thus when wrong, vision can produce *auditory* illusions.

As we saw from the Ames (1946) demonstration of the queer shaped room, the wrong perceptual hypothesis can cause basic sensory information to be wrongly interpreted and then we are misled by the most direct experience. The human seeing machine works very largely by means of a running hypothesis—which to us is reality though always it may be wrong. We are safeguarded from forced inactivity when sensory information is removed, by the prevailing hypothesis flywheeling along in the absence of correction by new data. It will gradually—though in unfortunate cases all too rapidly—get out of date as reality changes, but until then it is a great deal better than nothing. Seeing machines are not stimulus-response devices, but continuously relate incoming data to internal file cards which are kept up to date and serve to give meaning to the fleeting pictures in the eyes. They give due warning of the future, making possible elaborate strategies going far beyond mere reflex behaviour, which is but a hand-to-mouth existence. Thanks to the ancient survival value of space perception—which gives warning of the imminence of distant objects by signalling space and speed, together with identity of Friend and Foe—we now have in our brains seeing machines giving more detailed knowledge than we need for mere survival. Perceptual knowledge goes far beyond the purely sensory characteristics of objects, and is incorporated and becomes one with immediate experience. We see what we know—or perceptually 'think' we know.

Suppose that we wish to make a seeing machine to report an alien world. We must allow it to learn—to relate its past to its present—if it is to do more than transmit pictures which we must interpret with our brains, to do the seeing. If it is to do the seeing it must interpret its own images according to stored data; but these, from an alien world, we may not know. Communication demands a sea of shared assumptions and common data. An alien seeing machine is likely to be difficult to understand—for its filing cards will not be ours, and yet they are essential for it to interpret the pictures in its eyes—to see. It may learn how to survive, but the more surprising and interesting its information, the more difficult it will be for us to learn from its experience. It is tempting to build a seeing machine, and send it to distant worlds, but unfortunately its filing cards will either be inappropriate or give answers we will not understand—as the blind cannot ultimately understand those with sight though they use the 'same' words. The same is true of sending a human seeing machine—a man—to an alien environment. As he learns to see it, so will communication become more difficult and ultimately be impossible. If he, or it, does not learn to see the alien world (by failing to write new file cards to develop perceptual hypotheses) then the sensory information will be inappropriate or have no meaning. Communication will seem easy but the messages will be false. The alien seeing machine will be

GREGORY

riddled by illusions, through misinterpretation of the ambiguous images in its metal and glass eyes.

REFERENCES

- Ames, A. (1946). *Some demonstrations concerned with the origin and nature of our sensations*. Laboratory Manual. Dartmouth Eye Institute.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Held, R., & Hein, A. (1963). Movement produced stimulation in the development of visually guided behavior. *J. comp. Psychol.*, 56, 872-876.
- Helmholtz, H. von. (1925). *Physiological Optics*. J. P. Southall, Trs. New York: Optical Soc. Am.
- Lettvin, J. Y., Maturana, H. R., McCulloch, W. A., & Pitts, W. H. (1959). 'What the frog's eye tells the frog's brain.' *Proc. Inst. Radio Engrs*, 47, 1940-1951.
- Rushton, W. H. (1953). *Aspects of retinal physiology*. Rio de Janeiro: Instituto de Biofisica.
- Walls, G. L. (1963). *The vertebrate eye*. London: Hafner.