References


Loewenstein, W. R. (1960) Biological transducers. Scientific American, August (Also in Perception: Mechanisms and Models, Readings from Scientific American (1972), San Francisco: Freeman.)


Loewenstein, W. R. (1960) Biological transducers. Scientific American, August (Also in Perception: Mechanisms and Models, Readings from Scientific American (1972), San Francisco: Freeman.)


Understanding Complex Information-processing Systems

[... ] Almost never can a complex system of any kind be understood as a simple extrapolation from the properties of its elementary components. Consider, for example, some gas in a bottle. A description of thermodynamic effects—temperature, pressure, density, and the relationships among these factors—is not formulated by using a large set of equations, one for each of the particles involved. Such effects are described at their own level, that of an enormous collection of particles; the effort is to show that in principle the microscopic and macroscopic descriptions are consistent with one another. If one hopes to achieve a full understanding of a system as complicated as a nervous system, a develop- ing embryo, a set of metabolic pathways, a bottle of gas, or even a large computer program, then one must be prepared to contemplate different kinds of explanation at different levels of description that are linked, at least in principle, into a cohesive whole, even if linking the levels in complete detail is impractical. For the specific case of a system that solves an information-processing problem, there are in addition the twin strands of process and representation, and both these ideas need some discussion.

A musical score provides a way of representing a symphony; the alphabet allows the construction of a written representation of words; and so forth. The phrase “formal scheme” is critical to the definition, but the reader should not be frightened by it. The reason is simply that we are dealing with information-processing machines, and the way such machines work is by using symbols to stand for things—to represent things, in our terminology. To say that something is a formal scheme means only that it is a set of symbols with rules for putting them together—no more and no less.

A representation, therefore, is not a foreign idea at all—we all use representations all the time. However, the notion that one can capture some aspect of reality by making a description of it using a symbolic scheme and that to do so can be useful seems to be a fascinating and powerful idea. But even the simple examples we have discussed introduce some rather general and important issues that arise whenever one chooses to use one particular representation. For example, if one chooses the Arabic numeral representation, it is easy to discover whether a number is a power of 10 but difficult to discover whether it is a power of 2. If one chooses the binary representation, the situation is reversed. Thus, there is a trade-off; any particular representation makes certain information explicit at the expense of information that is pushed into the background and may be quite hard to recover.

This issue is important, because how information is represented can greatly affect how easy it is to do different things with it. This is evident even from our numbers example: It is easy to add, to subtract, and to even to multiply if the Arabic or binary representations are used, but it is not at all easy to do these things—especially multiplication—with Roman numerals. This is a key reason why the Roman culture failed to develop mathematics in the way the earlier Arabic cultures had.

An analogous problem faces computer engineers today. Electronic technology is much more suited to a binary number system than to the conventional base 10 system, yet humans supply their data and require the results in base 10. The design decision facing the engineer, therefore, is: Should one pay the cost of conversion into base 2, carry out the arithmetic in a binary representation, and then convert back into decimal numbers on output; or should one sacrifice efficiency of circuitry to carry out operations directly in a decimal representation? On the whole, business computers and pocket calculators take the second approach, and general purpose computers the first. But even though one is not restricted to using just one representation system for a given type of information, the choice of which to use is important and cannot be taken lightly. It determines what information is made explicit and hence what is pushed further into the background, and it has a far-reaching effect on the ease and difficulty of the operations that may subsequently be carried out on that information.

Process

The term process is very broad. For example, addition is a process, and so is taking a Fourier transform. But so is making a cup of tea, or going shopping. For the purposes of this book, I want to restrict our attention to the meanings associated with machines that are carrying out information-processing tasks. So let us examine in depth the notions behind one simple such device, a cash register at the checkout counter of a supermarket.

There are several levels at which one needs to understand such a device, and it is perhaps most useful to think in terms of three of them. The most abstract is the level of what the device does and why. What it does is arithmetic, so our first task is to master the theory of addition. Addition is a mapping, usually denoted by +, from pairs of numbers into single numbers, for example, + maps the pair (3, 4) to 7, and I shall write this in the form (3 + 4) → 7. Addition has a number of abstract properties, however. It is commutative: both (3 + 4) and (4 + 3) are equal to 7; and associative: the sum of 3 + (4 + 5) is the same as the sum of (3 + 4) + 5. Then there is the unique distinguished element, zero, the adding of which has no effect: (4 + 0) = 4. Also, there is the unique “inverse,” written (−4) inprivate for the algorithm we could choose the Arabic numerals for the representations, and for the algorithm we could follow the usual rules about adding the least significant digits first and “carrying” if the sum exceeds 9. Cash registers, whether mechanical or electronic, usually use this type of representation and algorithm.

There are three important points here. First, there is usually a wide choice of representation. Second, the choice of the device for carrying out one may depend rather critically on the particular representation that is employed. And third, even for a given fixed representation, there are often several possible algorithms for carrying out the same process. Which one is chosen will usually depend on any particularly desirable or undesirable characteristics that the algorithms may have. For example, one algorithm may be much more efficient than another, or there may be slightly less efficient but more robust (that is, less sensitive to slight inaccuracies in the data on which it must run). Or again, one algorithm may be parallel, and another serial. The choice, then, may depend on the type of hardware or machinery in which the algorithm is to be embodied physically.

This brings us to the third level, that of the device in which the process is to be realized physically. The important point here is that, once again, the same algorithm may be implemented in quite different technologies. The child who methodically adds two numbers from right to left, carrying a digit when necessary, may be using the same algorithm that is implemented by the wires and transistors of the cash register in the neighborhood supermarket, but the physical realization of the algorithm is quite different in these two cases. Another example: Many people have written computer programs to play tic-tac-toe, and there is a more or less standard algorithm that cannot lose. This algorithm has in fact been implemented by W. D. Hills and B. Silverman in a quite different technology, in a computer made out of Tinkertoys, a children’s wooden building set. The whole monstrously ungainly engine, which actually works, currently resides in a museum at the University of Missouri in St. Louis.

Some styles of algorithm will suit some physical substrates better than others. For example, in conventional digital computers, the number of connections is comparable to the number of gates, while in a brain, the number of connections is
much larger \(\times 10^3\) than the number of nerve cells. The underlying reason is that wires are rather cheap in biological architecture, because they can grow individually and in three dimensions. In conventional technology, wire laying is more or less restricted to two dimensions, which quite severely restricts the scope for using parallel techniques and algorithms; the same operations are often better carried out serially.

The three levels

We can summarize our discussion in something like the manner shown in table 5.1, which illustrates the different levels at which an information-processing device must be understood before one can be said to have understood it completely. At one extreme, the top level, is the abstract computational theory of the device, in which the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated. In the center is the choice of representation for the input and output and the algorithm to be used to transform one into the other. And at the other extreme are the details of how the algorithm and representation are realized physically—the detailed computer architecture, so to speak. These three levels are coupled, but only loosely. The choice of an algorithm is influenced for example, by what it has to do and by the hardware in which it must run. But there is a wide choice available at each level, and the explication of each level involves issues that are rather independent of the other two.

Each of the three levels of description will have its place in the eventual understanding of perceptual information processing, and of course they are logically and causally related. But an important point to note is that since the three levels are only rather loosely related, some phenomena may be explained at only one or two of them. This means, for example, that a correct explanation of some psychophysical observation must be formulated at the appropriate level. In attempts to relate psychophysical problems to physiology, too often there is confusion about the level at which problems should be addressed. For instance, some are related mainly to the physical mechanisms of vision—such as afterimages (for example, the one you see after staring at a light bulb) or such as the fact that any color can be matched by a suitable mixture of the three primaries (a consequence principally of the fact that we humans have three types of cones). On the other hand, the ambiguity of the Necker cube (figure 5.1) seems to demand a different kind of explanation. To be sure, part of the explanation of its reversal must have to do with a bistable neural network (that is, one with two distinct stable states) somewhere inside the brain, but few would feel satisfied by an account that failed to mention the existence of two different but perfectly plausible three-dimensional interpretations of this two-dimensional image.

For some phenomena, the type of explanation required is fairly obvious. Neuroanatomy, for example, is clearly tied principally to the third level, the physical realization of the computation. The same holds for synaptic mechanisms, action potentials, inhibitory interactions, and so forth. Neurophysiology, too, is related mostly to this level, but it can also help us to understand the type of representations being used, particularly if one accepts something along the lines of Barlow’s views that I quoted earlier. But one has to exercise extreme caution in making inferences from neurophysiological findings about the algorithms and representations being used, particularly until one has a clear idea about what information needs to be represented and what processes need to be implemented.

Psychophysics, on the other hand, is related more directly to the level of algorithm and representation. Different algorithms tend to fail in radically different ways as they are pushed to the limits of their performance or are deprived of critical information. As we shall see, primarily psychophysical evidence proved to Poggio and myself that our first stereo-matching algorithm (Marr and Poggio, 1976) was not the one that is used by the brain, and the best evidence that our second algorithm (Marr and Poggio, 1979) is roughly the one that is used also comes from psychophysics. Of course, the underlying computational theory remained the same in both cases, only the algorithms were different.

Psychophysics can also help to determine the nature of a representation. The work of Roger Shepard (1975), Eleazar Rosch (1978), or Elizbeth Warrington (1975) provides some interesting hints in this direction. More specifically, Stevens (1979) argued from psychophysical experiments that surface orientation is represented by the coordinates of slant and tilt, rather than (for example) the more traditional \((\rho, \phi)\) of gradient space (see chapter 3). He also deduced from the uniformity of the size of errors made by subjects judging surface orientation over a wide range of orientations that the representational quantities used for slant and tilt are pure angles and not, for example, their cosines, sines, or tangents.

More generally, if the idea that different phenomena need to be explained at different levels is kept clearly in mind, it often helps in the assessment of the validity of the different kinds of objections that are raised from time to time. For example, one favorite is that the brain is quite different from a computer because one is parallel and the other serial. The answer to this, of course, is that the distinction between serial and parallel is a distinction at the level of algorithm; it is not fundamental at all—anything programmed in parallel can be rewritten serially (though not necessarily vice versa). The distinction, therefore, provides no grounds for arguing that the brain operates so differently from a computer that a computer could not be programmed to perform the same tasks.

Importance of computational theory

Although algorithms and mechanisms are empirically more accessible, it is the top level, the level of computational theory, which is critically important from an information-processing point of view. The reason for this is that the nature of the computations that underlie perception depends more upon the computational problems that have to be solved than upon the particular hardware in which their solutions are implemented. To phrase the matter another way, an algorithm is likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied.

In a similar vein, trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers. It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds’ wings make sense. More to the point, as we shall see, we cannot understand why retinal ganglion cells and lateral geniculate neurons have the receptive fields they do just by studying
their anatomy and physiology. We can understand how these cells and neurons behave as they do by studying their wiring and interactions, but in order to understand their functional properties as they are - which are they are circularly symmetrical and why their excitatory and inhibitory regions have characteristic shapes and distributions - we have to know a little of the theory of differential operators, studying their wiring and interactions, but in order how these cells and neurons behave as they do by the mathematics of the uncertainty principle (see chapter 2).

Perhaps it is surprising that the very specialized empirical disciplines of the neurosciences failed to appreciate fully the absence of computational theory, but it is surprising that this level of approach did not play a more forceful role in the early development of artificial intelligence. For far too long, a heuristic program for carrying some task was held to be a theory of the task, and the distinction between what a program did and how it did it was not taken seriously. As a result, (1) a style of explanation evolved that invoked the use of special mechanisms to solve particular problems, (2) particular data structures, such as the lists of attribute value pairs called property lists in the LISP programming language, were held to amount to understanding, and (3) there was frequently no way to determine whether a program would deal with a particular case other than by running the program.

Failure to recognize this theoretical distinction between what and how also greatly hampered communication between the fields of artificial intelligence and linguistic theory. Chomsky's (1965) theory of transformational grammar is a true computational theory in the sense defined earlier. It is concerned solely with specifying what the syntactic decomposition of an English sentence should be, and not at all with how that decomposition should be achieved. Chomsky himself was very clear about this - it is roughly his distinction between competence and performance, though his idea of performance did include other factors, like stopping in midutterance - but the fact that his theory was not at all concerned with the function of the brain was to "detect invariants" despite changes in "sensations" of light. This led him to a view in which the function of the brain was to "decode signals, not to interpret messages, nor to accept images, nor to organize the sensory input or to process the data, in modern terminology. It is to seek and extract information about the environment from the flow of ambient energy." He says that 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 the data, in modern terminology. It is to seek and extract information about the environment from the flow of ambient energy." He says that 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 the data, in modern terminology. It is to seek and extract information about the environment from the flow of ambient energy." He says that 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 the data, in modern terminology. It is to seek and extract information about the environment from the flow of ambient energy."

The approach of J. J. Gibson

In perception, perhaps the nearest anyone came to the level of computational theory was Gibson (1966). However, although some aspects of his thinking were on the right lines, he did not understand properly what information processing was, which led him to seriously underestimate the complexity of the information-processing problems involved in vision and the consequent subtility that is necessary in approaching them.

Gibson's important contribution was to take the debate away from the philosophical considerations of sense-data and the affective qualities of sensations and to note instead that the important thing about the senses is that they are channels for perception of the real world outside of, in the case of vision, of the visible surfaces. He therefore asked the critically important question: How does one obtain constant perceptions of an everyday life on the basis of continually changing sensations? This is exactly the right question, showing that Gibson correctly regarded the problem of perception as that of recovering from sensory information "valid" properties of the external world. His problem was that he had a much over-simplified view of how this should be done. His approach led him to consider higher-order variables - stimulus energy, ratios, proportions, and so on - as "invariants" of the movement of an observer and of changes in stimulation intensity.

"These invariants," he wrote, "correspond to permanent properties of the environment. They constitute, therefore, information about the permanent environment." This led him to a view in which the function of the brain was to "detect invariants" despite changes in "sensations" of light. This was the basic idea behind the notion of ecological optics (Gibson, 1966, 1979). Although one can criticize certain shortcomings in the quality of Gibson's analysis, its major and, in my view, fatal shortcoming lies at a deeper level and results from a failure to realize two things. First, the detection of physical invariants, like image surfaces, is exactly and precisely an information-processing problem, in modern terminology. And second, he vastly underrated the sheer difficulty of such detection. In discussing the recovery of three-dimensional information from the movement of an observer, he says that "in motion, perspective information alone can be used" (Gibson, 1966: 202). And perhaps the key to Gibson is the following:

"The detection of non-change when an object moves in the world is not as difficult as it might appear. It is only made to seem difficult when we assume that the perception of constant dimensions of the object must depend on the correcting of sensations of inconstant form and size. The information for the constant dimension of an object is normally carried by invariant relations in an optic array Rigidity is specified (emphasis added)"

Yes, to be sure, but how? Detecting physical invariants is as difficult as Gibson feared, but nevertheless we do it. And the only way to understand how is to treat it as an information-processing problem.

The underlying point is that visual information processing is actually very complicated, and Gibson was not the only thinker who was misled by the apparent simplicity of the act of seeing. The whole tradition of philosophical inquiry into the nature of perception seems not to have taken seriously enough the complexity of the information-processing problems involved. Austin's (1962) Sense and Sensibility entertainingly demolishes the argument, apparently favored by earlier philosophers, that since we are sometimes deluded by illusions (for example, a straight stick appears bent if it is partly submerged in water), we see nonsense rather than material things. The answer is simply that usually our perceptual processing does run correctly (it delivers a true description of what is there), but although evolution has seen to it that our processing allows for many changes (like inconstant illumination), the perturbation due to the refraction of light by water is not one of them. And incidentally, although the example of the bent stick has been discussed since Aristotle, I have seen no philosophical inquiry into the nature of the perceptions of, for instance, a hero, which is a bird that feeds by pecking up fish first seen from above the water surface. For such birds the visual correction might be present.

Anyway, my main point here is another Austin's (1962) view of time on the idea that perception tells one not about real properties of the external world, and one thing he considers is "real shape" (p. 66), a notion which had cropped up earlier in his discussion of a coin that "looked elliptical" from some points of view. Even so, it had a real shape which remained unchanged. But coins in fact are rather special cases. For one thing their outlines are well defined and very highly stable, and for another they have a known and a nameable shape. But there are plenty of things of which this is not true. What is the real shape of a cloud? or of a cat? Does its real shape change when we take a picture of it, or, alternatively, how do we know that it does not change from moment to moment? What posture is its real shape on display? Furthermore, is its real shape such as to be fairly smooth outlines, or must it be finely enough serrated to take account of each hair? It is pretty obvious that there is no answer to these questions - no rules according to which, no procedure by which, answers are to be determined (emphasis added) (p. 67).

But there are answers to these questions. There are ways of describing the shape of a cat to an arbitrary level of precision (see chapter 5), and there are rules and procedures for arriving at
such descriptions. That is exactly what vision is about, and precisely what makes it complicated.

A Representational Framework for Vision

Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information (Marr, 1976; Marr and Nishihara, 1978). We have already seen that a process may be thought of as a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it consists of arrays of image intensity values as detected by the photoreceptors in the retina.

It is quite proper to think of an image as a representation: the items that are made explicit are the image intensity values at each point in the array, which we can conveniently denote by \( I(x,y) \) at coordinate \((x,y)\). In order to simplify our discussion, we shall neglect for the moment the fact that there are several different types of receptor, and imagine instead that there is just one, so that the image is black-and-white. Each value of \( I(x,y) \) thus specifies a particular level of gray; we shall refer to each detector as a picture element or pixel to the whole array \( I \) as an image.

But what of the output of the process of vision? We have already agreed that it must consist of a useful description of the world, but that requirement is rather ill-defined. Can we go better? Well, it is perfectly true that, unlike the input, the result of vision is much harder to discern, let alone specify precisely, and an important aspect of this new approach is that it makes quite concrete proposals about what that end is. But before we begin that discussion, let us step back a little and spend a little time formulating the more general issues that are raised by these questions.

The purpose of vision

The usefulness of a representation depends upon how well suited it is to the purpose for which it is used. A pigeon uses vision to help it navigate, fly, and seek out food; many types of jumping spider use vision to tell the difference between a potential mate and a potential meal. One type, for example, has a curious retina composed of two diagonal strips arranged in a V. If it detects a red V on the back of an object lying in front of it, the spider has found a mate. Otherwise, maybe a meal. The frog detects bugs with its retina; and the rabbit retina is full of special gadgets, including what is apparently a hawk detector, since it responds well to the pattern made by a prayer hawk hovering overhead. Human vision, on the other hand, seems to be very much more general, although it clearly contains a variety of special-purpose mechanisms that can, for example, direct the eye toward an unexpected movement in the visual field or cause one to blink or otherwise avoid something that approaches one's head too quickly.

Vision, in short, is used in such a bewildering variety of ways that the visual systems of different animals must differ significantly from one another. Can the type of formulation that I have been advocating, in terms of representations and processes, possibly prove adequate for them all? I think so. The general point here is that because vision is used by different animals for such a wide variety of purposes, it is inconceivable that all the animals use the same representations; each can confidently be expected to use one or more representations that are nicely tailored to the owner's purposes.

As an example, let us consider briefly a primitive but highly efficient visual system that has the added virtue of being well understood: Werner Reichardt's group in Tübingen has spent the last 14 years patiently unraveling the visual flight-control system of the housefly, and in a famous collaboration, Reichardt and Tomaso Poggio have gone far toward solving the problem (Reichardt and Poggio, 1976, 1979; Poggio and Reichardt, 1976). Roughly speaking, the fly's visual apparatus controls its flight through a collection of about five independent, rigidly inflexible, very fast responding systems (the time from visual stimulus to change of torque is only 21 ms). For example, one of these systems is the landing system; if the visual field "explodes" fast enough (because a surface looms nearby), the fly automatically inverts to land (the torque produced by the asymmetry of the horizontal thrust from the left and right wings). The visual input to the horizontal control system, for example, is completely described by the two terms

\[ r(\psi) \psi + D(\psi) \]

where \( r \) and \( D \) have the form illustrated in figure 5.2. This input describes how the fly tracks an object that is present at angle \( \psi \) in the visual field and has angular velocity \( \psi \). This system is triggered to track objects of a certain angular dimension in the visual field, and the main strategy is such that if the visible object was another fly a few inches away, then it would be intercepted successfully. If the target was an elephant 100 yd away, interception would fail because the fly's built-in parameters are for another fly nearby, not an elephant far away.

Thus, fly vision delivers a representation in which at least these three things are specified: (1) whether the visual field is looming sufficiently fast that the fly should contemplate landing; (2) whether there is a small patch—it could be a black speck or, it turns out, a textured figure in front of a textured ground—having some kind of motion relative to its background; and if there is such a patch, (3) \( \psi \) and \( \psi \) for this patch are delivered to the motor system. And that is probably about 60% of fly vision. In particular, it is extremely unlikely that the fly has any explicit representation of the visual world around him—no true conception of a surface, for example, but just a few triggers and some specifically fly-centered parameters like \( \psi \) and \( \psi \).

It is clear that human vision is much more complex than this, although it may well incorporate subsystems not unlike the fly's to help with specific and rather low-level tasks like the control of pursuit eye movements. Nevertheless, as Poggio and Reichardt have shown, even these simple systems can be understood in the same sort of way, as information-processing tasks. And one of the fascinating aspects of their work is how they have managed not only to formulate the differential equations that accurately describe the visual control system of the fly but also to express these equations, using the Volterra series expansion, in a way that gives direct information about the minimum possible complexity of connections of the underlying neuronal networks.
Advanced vision

Visual systems like the fly's serve adequately and with speed and precision the needs of their owners, but they are not very complicated; very little objective information about the world is obtained. The information is all very much subjective — objective information about the world is obtained but they are not very complicated; very little angular size of the stimulus as the fly sees it rather than the objective size of the object out there, the angle that the object has in the fly's visual field rather than its position relative to the fly or to some external reference, and the object's angular velocity, again in the fly's visual field, rather than any assessment of its true velocity relative to the fly or to some stationary reference point.

One reason for this simplicity must be that these facts provide the fly with sufficient information for it to survive. Of course, the information is not optimal and from time to time the fly will flitter away its energy chasing a falling leaf a medium distance away or an elephant a long way away as a direct consequence of the inadequacies of its perceptual system. But this apparently does not matter very much — the fly has sufficient excess energy for it to be able to absorb these extra costs.

Another reason is certainly that translating these rather subjective measurements into more objective qualities involves much more computation. How, then, should one think about more advanced visual systems — human vision, for example. What are the issues? What kind of information is vision really delivering, and what are the representational issues involved?

My approach to these problems was very much influenced by the fascinating accounts of clinical neurology, such as Critchley (1953) and Warrington and Taylor (1973). Particularly important was a lecture that Elizabeth Warrington gave at MIT in October 1973, in which she described the capacities and limitations of patients who had suffered left or right parietal lesions. For me, the most important thing that she did was to draw a distinction between the two classes of patient (see Warrington and Taylor, 1978). For those with lesions on the right side, recognition of a common object was possible provided that the patient's view of it was in some sense straightforward. She used the words conventional and unconventional — a water pail or a clarinet seen from the side gave "conventional" views but seen end-on gave "unconventional" views. If these patients recognized the object at all, they knew its name and its semantics — that is, its use and purpose, how big it was, how much it weighed, what it was made of, and so forth. If their view was unconventional — a pail seen from above, for example — would the patients fail to recognize it, but they would vehemently deny that it could be a view of a pail. Patients with left parietal lesions behaved completely differently. Often these patients had no language, so they were unable to name the viewed object or state its purpose and semantics. But they could convey that they correctly perceived its geometry, its shape — even from the unconventional view.

Warrington's talk suggested two things. First, the representation of the shape of an object is stored in a different place and is therefore a quite different kind of thing from the representation of its use and purpose. And second, vision alone can deliver an internal description of the shape of a viewed object, even when the object was not recognized in the conventional sense of understanding its use and purpose.

This was an important moment for me for two reasons. The general trend in the computer vision community was to believe that recognition was so difficult that it required every possible kind of information. The results of view 2-Appearly appeared a few years later in programs like Freuder's (1974) and Tenenbaum and Barrow's (1976). In the latter program, knowledge about offices — for example, that desks have telephones on them and that telephones are black — was used to help "segment" out a black blob halfway up an image and "recognize" it as a telephone. Freuder's program used a similar approach to "segment" and "recognize" a hammer in a scene. Clearly, we use such knowledge in real life; I once saw a brown blob quivering amongst the lettuce in my garden and correctly identified it as a rabbit, even though the visual information alone was inadequate. And yet here was this young woman calmly telling us not only that her patients could convey to her that they had grasped the shapes of things that she had shown them, even though they could not name the objects or say how they were used, but also that they could happily continue to do so even if she had made the task extremely difficult visually by showing them peculiar views or by illuminating the objects in peculiar ways. It seemed clear that the intuitions of the computer vision people were completely wrong and that even in difficult circumstances shapes could be determined by vision alone.

The second important thing, I thought, was that Elizabeth Warrington had put her finger on what was somehow the quintessential fact of human vision — that it tells us about shape and space and spatial arrangement. Here lay a way to formulate its purpose — building a description of the shapes and positions of things from images. Of course, this is by no means all that vision can do; it also tells about the illumination and about the reflectances of the surfaces that make the shapes — their brightness and colors and visual textures — and about their motion. But these things seemed secondary; they could be hung off a theory in which the main job of vision was to derive a representation of shape.

To the desirable via the possible

Finally, one has to come to terms with cold reality. Desirable as it may be to have vision deliver a completely invariant shape description from an image (whatever that may mean in detail), it is almost certainly impossible in only one step. We can only do what is possible and proceed from there toward what is desirable. Thus we arrived at the idea of a sequence of representations, starting with the representation that can be obtained straight from an image but that are carefully designed to facilitate the subsequent recovery of gradually more objective, visually properties about an object's shape. The main stepping stone toward this goal is describing the geometry of the visible surfaces, since the information encoded in images, for example by stereo, shading, texture, contours, or visual motion, is due to a shape's local surface properties. The objective of many early visual computations is to extract this information.

However, this description of the visible surfaces turns out to be unsuitable for recognition tasks. There are several reasons why, perhaps the most prominent being that like all early visual processes, it depends critically on the viewpoint. The final step therefore consists of transforming the viewer-centered surface description into a representation of the three-dimensional shape and spatial arrangement of an object that does not depend upon the direction from which the object is being viewed. This final description is object centered rather than viewer centered.

The overall framework described here therefore divides the derivation of shape information from images into three representational stages (table 5.2): (1) the representation of properties of the two-dimensional image, such as intensity changes and local two-dimensional geometry; (2) the representation of properties of the visible surfaces in a viewer-centered coordinate system, such as surface orientation, distance from the viewer, and discontinuities in these quantities; surface reflectance; and some coarse description of the prevailing illumination; and (3) an object-centered representation of the three-dimensional structure and of the organization of the viewed shape, together with some description of its surface properties.
This framework is summarized in Table 5.2. Chapters 2 through 5 give a more detailed account.

**Synopsis**

Our survey of this new, computational approach to vision is now complete. Although there are many gaps and omissions, I hope that it is solid enough to establish a firm point of view about the subject and to prompt the reader to begin to judge its value. In this brief chapter, I shall take a very broad view of the whole approach, inquiring into its most important general features and how they relate to one another, and trying to say something about the style of research that this approach implies. It is convenient to divide the discussion into four main points.

The first point is one that we have met throughout the account—the notion of different levels of explanation. The central tenet of the approach is that to understand what vision is and how it works, one has to do all these things at once and also be very aware of the additional level of explanation that I have called the level of computational theory. The recognition of the existence and importance of this level is one of the most important aspects of this approach. Having recognized this, one can formulate the three levels of explanation explicitly (computational theory, algorithm, and implementation), and it then becomes clear how these different levels are related to the different representations and processes, which is set out explicitly in figure 5.3, often provides a useful aid to thinking how best to proceed when studying computational theory are valid and may be implicit in the human processor; second, if a process matches human performance, it is probably sufficiently powerful to form part of a general purpose vision machine.

The final point concerns the methodology or style of this type of approach, and it involves two main observations. First, the duality between representations and processes, which is set out explicitly in figure 5.3, often provides a useful aid to thinking how best to proceed when studying a particular problem. In the study both of representations and of processes, general problems are often suggested by everyday experience or by psychophysical or even neurophysiological findings of a quite general nature. Such general observations can often lead to the formulation of a particular process or representational theory, specific examples of which can be programmed or subjected to detailed psychophysical testing. Once we have sufficient confidence in the correctness of the process or representation at this level, we can inquire about its detailed implementation, which involves the ultimate and very difficult problems of neurophysiology and neuroanatomy.

The second observation is that there is no real recipe for this type of research—even though I have sometimes suggested that there is—any more than there is a straightforward procedure for discovering things in any other branch of science. Indeed, part of the fun is that we never really know where the next key is going to come from—a piece of daily experience, the report of a neurological deficit, a theorem about three-dimensional geometry, a psychophysical finding in hyperacuity, a neurophysiological observation, or the careful analysis of a representational problem. All these kinds of information have played important roles in establishing the framework that I have described, and they will presumably continue to contribute to its advancement in an interesting and unpredictable way. I hope only that these observations may persuade some of my readers to join in the adventures we have had and to help in the long but rewarding task of unraveling the mysteries of human visual perception.

### Table 5.2

<table>
<thead>
<tr>
<th>Representational problem</th>
<th>Computational problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of information to be made explicit</td>
<td>Computational theory (processes and constraints)</td>
</tr>
<tr>
<td>Specific representation (can be programmed)</td>
<td>Specific algorithm (can be programmed)</td>
</tr>
<tr>
<td>Detailed psychophysics</td>
<td>Detailed neurophysiology and neuroanatomy</td>
</tr>
</tbody>
</table>

### Figure 5.3

Relationships between representations and processes

---

Note

1. Editor’s note: the passages to which Marr here refers are as follows (from pp 12-13 of *Vision*).

If one explores the responsiveness of single ganglion cells in the frog's retina using handheld targets, one finds that one particular type of ganglion cell is most effectively driven by something like a black disc subtending a degree or so moved rapidly to and fro within the unit's receptive field. This causes a vigorous discharge which can be maintained without much decrement as long as the movement is continued. Now, if the stimulus which is optimal for this class of cells is presented to intact frogs, the behavioral response is often dramatic; they turn towards the target and make repeated feeding responses consisting of a jump and snap. The selectivity of the retinal neurons and the frog's reaction when they are selectively stimulated, suggest that they are "bug detectors" (Barlow, 1953) performing a primitive but vitally important form of recognition. The result makes one suddenly realize that a large part of the sensory machinery involved in a frog's feeding responses may actually reside in the retina rather than in mysterious "centers" that would be too difficult to understand by physiological methods. The essential lock-and-key property resides in each member...
of a whole class of neurons and allows the cell to discharge only to the appropriate key pattern of sensory stimulation. Lettvin et al. (1959) suggested that there were five different classes of cell in the frog, and Barlow, Hill, and Levick (1964) found an even larger number of categories in the rabbit (Barlow et al.) called these key patterns "trigger features," and Maturana et al. (1960) emphasized another important aspect of the behavior of these ganglion cells: a cell continues to respond to the same trigger feature in spite of changes in light intensity over many decades. The properties of the retinas are such that a ganglion cell can, figuratively speaking, reach out and determine that something specific is happening in front of the eye. Light is the agent by which it does this, but it is the detailed pattern of the light that carries the information, and the overall level of illumination prevailing at the time is almost totally disregarded (Barlow, 1972: 37).

The cumulative effect of all the changes I have tried to outline above has been to make us realize that each single neuron can perform a much more complex and subtle task than had previously been thought [emphasis added]. Neurons do not loosely and unreliably remap the luminous intensities of the visual image onto our sensorium, but instead they detect pattern elements, discriminate the depth of objects, ignore irrelevant causes of variation and are arranged in an intriguing hierarchy. Furthermore, there is evidence that they give prominence to what is informationally important, can respond with great reliability, and can have their pattern selectivity permanently modified by early visual experience. This amounts to a revolution in our outlook. It is now quite inappropriate to regard unit activity as a noisy indication of more basic and reliable processes involved in mental operations: instead, we must regard single neurons as the prime movers of these mechanisms. Thinking is brought about by neurons and we should not use phrases like "unit activity reflects, reveals or monitors thought processes," because the activities of neurons, quite simply, are thought processes.

This revolution stemmed from physiological work and makes us realize that the activity of each single neuron may play a significant role in perception (ibid.: 380).

References

Each volume in this outstanding series provides an authoritative and comprehensive collection of the essential primary readings from philosophy's main fields of study. Designed to complement the Blackwell Companions to Philosophy series, each volume represents an unparalleled resource in its own right, and will provide the ideal platform for course use.

1. Cottingham: Western Philosophy: An Anthology
2. Cahoon: From Modernism to Postmodernism: An Anthology
3. LaFollette: Ethics in Practice: An Anthology
4. Goodin and Pettit: Contemporary Political Philosophy: An Anthology
5. Eze: African Philosophy: An Anthology
7. Kim and Sosa: Metaphysics: An Anthology
11. Sosa and Kim: Epistemology: An Anthology

Edited by

Robert Cummins
University of California at Davis

and

Denise Dellarosa Cummins
University of California at Davis