Knowledge-based vision and simple visual machines

DAVE CLIFF AND JASON NOBLE

School of Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QH, UK
(davec@cogs.susx.ac.uk) (jasonn@cogs.susx.ac.uk)

SUMMARY

The vast majority of work in machine vision emphasizes the representation of perceived objects and events: it is these internal representations that incorporate the 'knowledge' in knowledge-based vision or form the 'models' in model-based vision. In this paper, we discuss simple machine vision systems developed by artificial evolution rather than traditional engineering design techniques, and note that the task of identifying internal representations within such systems is made difficult by the lack of an operational definition of representation at the causal mechanistic level. Consequently, we question the nature and indeed the existence of representations posited to be used within natural vision systems (i.e. animals). We conclude that representations argued for on *a priori* grounds by external observers of a particular vision system may well be illusory, and are at best place-holders for yet-to-be-identified causal mechanistic interactions. That is, applying the knowledge-based vision approach in the understanding of evolved systems (machines or animals) may well lead to theories and models that are internally consistent, computationally plausible, and entirely wrong.

1. INTRODUCTION

The vast majority of work in machine vision emphasizes the representation of perceived objects and events: it is these internal representations that are the 'knowledge' in knowledge-based vision and the 'models' in model-based vision. In this paper, we argue that such notions of representation may have little use in explaining the operation of simple machine vision systems that have been developed by artificial evolution rather than through traditional engineering design techniques, and which are, therefore, of questionable value in furthering our understanding of vision in animals, which are also the product of evolutionary processes.

This is not to say that representations do not exist or are not useful: there are many potential applications of machine vision, of practical engineering importance, where significant problems are alleviated or avoided altogether by use of appropriate structured representations. Examples include medical imaging, terrain mapping, and traffic monitoring (e.g. Taylor *et al.* 1986; Sullivan 1992).

But the success of these engineering endeavours may encourage us to assume that similar representations are of use in explaining vision in animals. In this paper, we argue that such assumptions may be misleading. Yet the assumption that vision is fundamentally dependent on representations (and further assumptions involving the nature of those representations) is widespread. We seek only to highlight problems with these assumptions; problems which appear to stem from incautious use of the notion of 'representation'. We argue in particular

that the notion of representation as the construction of an internal model representing some external situation is probably not applicable to evolved systems. This paper is intentionally provocative; the arguments put forward below are offered for discussion, rather than as unquestionable truths.

We start, in § 2, by briefly reviewing two key influences in the development of the view of vision as a process that forms representations for subsequent manipulation. Then, in § 3, we discuss simple visual machines by (i) summarizing the process of artificial evolution, (ii) then reviewing work where artificial evolution has been used to evolve design specifications for visual sensorimotor controllers, and (iii) discussing the issue of identifying representations in these evolved designs. Following this, § 4 explores further the issue of defining the notion of representation with sufficient accuracy for it to be of use in empirically determining whether representations are employed by a system. Finally, in § 5 we explore the implications of these issues for the study of vision in animals, before offering our conclusions in § 6.

2. BACKGROUND

Although it is beyond the scope of this paper to provide a complete historical account of the key influences on the development of present knowledge-based vision techniques and practices, there are two major works that permeate almost all knowledge-based vision with which we are familiar. These are the Physical Symbol System Hypothesis of Newell & Simon (1976) and Marr's (1982) work on vision.

1165

(a) The Physical Symbol System hypothesis

Newell & Simon (1976) were instrumental in establishing the belief that systems which engage in the syntactic manipulation of symbols and symbol structures have the necessary and sufficient means for general intelligent action. For Newell & Simon the symbols are arbitrary, but their interpretation and semantics (i.e. what the symbols represent) are socially agreed between observers of the symbol system. Under this hypothesis, intelligent action involves the receipt of symbols from symbol-generating sensory apparatus, the subsequent manipulation of those symbols (e.g. by using techniques derived from mathematical logic, or algorithmic search), in order to produce an output symbol or symbol structure. Both the input and the output have meaning conferred on them by external observers, rather than the meaning being intrinsic to the symbol (Harnad 1990).

In the field of artificial intelligence, Newell & Simon's hypothesis licensed a paradigm of research concentrating on intelligence as the manipulation of symbolic representations, and on perception as the generation of those symbols and symbol structures. Specialized symbol-manipulating and logic-based computer programming languages such as Lisp (e.g. Winston & Horn 1980) and Prolog (e.g. Clocksin & Mellish 1984) (from 'LISt Processing' and 'PROgramming in LOGic', respectively) were developed to ease the creation of 'knowledge-based systems' (e.g. Gonzalez & Dankel 1993). In due course, undergraduate textbooks appeared that essentially treated the hypothesis as an axiomatic truth (e.g. Nilsson 1982; Charniak & McDermott 1985), paying little attention to criticisms of the approach (e.g. Dreyfus 1979, 1981).

In the field of machine vision, the Physical Symbol System Hypothesis underwrites all research on knowledge-based vision, where it is assumed that the aim of vision is to deliver symbolic representations (or 'models') of the objects in a visual scene: in the words of Pentland (1986), to go 'from pixels to predicates'. This mapping from visual images to predicate-level representations was studied in depth by David Marr.

(b) Marr's theories of vision

Marr's (1982) work on vision had an enormous impact on practices in machine vision. He argued forcefully and coherently for vision to be treated as a data-driven, bottom-up process which delivers representations of three-dimensional (3D) shape from two-dimensional (2D) images. Marr cites studies of vision in humans as being influential in the development of his theories: in particular the mental rotation experiments of Shepard & Metzler (1971) and the parietal lesion data of Warrington & Taylor (1973, 1978).

In Shepard & Metzler's experiments, human subjects were shown pairs of line-drawings of simple objects, and were asked to discriminate whether the two images were projections of the same 3D object viewed from different poses, or images of two different but mirror-symmetric objects viewed from different poses. Their results (which remain the subject of debate) indicated that the length of time taken for subjects to

identify that the two images differed only in pose (i.e. were of the same object) was linearly related to the degree of 3D rotation involved in the difference in pose. From these results (and, indeed, via introspection if one attempts to perform this discrimination task) it is compelling to conclude that the nervous system generates some internal representation of 3D shape from one 2D image, and then somehow manipulates it to determine whether it can match the second 2D image.

Warrington & Taylor's results concerned human patients who had suffered brain lesions in the left or right parietal areas. Left-lesioned patients could perceive the shape of an object from a wide variety of poses, but could offer little or no description of its 'semantics': its name or its purpose. Meanwhile, right-lesioned patients could describe the semantics of an object, provided it was presented from a 'conventional' pose or view-angle; if the view was somehow 'unconventional', such as a clarinet viewed end-on, the right-lesioned patients would not be able to recognize the object, and in some cases they would actively dispute that the view could be one of that object.

These results, and other considerations, led Marr to conclude that the main job of vision is to derive representations of the shapes and positions of things from images. Other issues (such as the illumination and reflectances of surfaces; their brightness and colours and textures; their motion) '... seemed secondary' (Marr 1982, p. 36).

In Marr's approach, vision is fundamentally an information-processing task, attempting to recover 3D information hidden or implicit in the 2D image. Marr proposed that such information-processing tasks, or the devices that execute them, should be analysed using a three-level methodology:

'[There are three] 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.' (Marr 1982, p. 24.)

Application of this three-level methodology to the problem of analysing vision led Marr and his colleagues to develop a theory of vision involving a pipeline of processes applying transformations to intermediate representations derived from the initial image (Marr 1982, p. 37): the ambient optic array is sampled to form a 2D image, which represents intensities; the image is then operated on to form the 'primal sketch', which represents important information about the 2D image such as the intensity changes and their geometrical distribution and organization. Following this, the primal sketch is processed to form the '2½D sketch', which represents orientation and rough depth

of visible surfaces, and any contours of discontinuities in these quantities, still in a viewer-centred coordinate frame. Next, the 2½D sketch is processed to form an internal '3D model', which represents shapes and their spatial organization in an object-centred coordinate frame; including information about volume. Hence, the 3D model is an internal reconstruction of the external physical world.

Within Marr's framework, formation of the 3D model is the end of the visual process, and the model is then passed to 'higher' processes, such as updating or matching against a stored library of 3D shapes. Since the initial development and publication of these ideas, much knowledge-based vision has been based on this approach.

Over the last decade, the increasing research activity in 'active vision' (e.g. Ballard 1991), where the camera that forms the image is under dynamic control of the vision system, has led to a number of criticisms being levelled at Marr's approach (e.g. Nelson 1991; Horswill 1993).

3. SIMPLE VISUAL MACHINES

Traditional modular engineering design techniques, based on dividing a given problem into a number of sub-problems such that each sub-problem can be resolved using a separate computational module, require intermediate representations for inter-module communication. The task of each computational module is to receive input data in a pre-specified representation, apply some required transformation, and pass on the result of the transformation as the output of the module. The Marr pipeline is a fine example of this approach: to go from image to 3D model in one step is unrealistically ambitious; instead, a sequence of operations is applied to the image, generating successive internal representations, leading to the final desired representation. Given that such techniques are well-established in engineering design and manifestly successful in a number of potentially very problematic task domains, it is difficult to conceive of alternatives.

However, recent work in adaptive behaviour (see the journal *Adaptive Behavior*, published by MIT Press, or the proceedings of the biennial conference on simulation of adaptive behaviour (Meyer & Wilson 1991; Meyer *et al.* 1993; Cliff *et al.* 1994; Maes *et al.* 1996)) has employed *artificial evolution* (i.e. genetic algorithms) as an alternative to traditional design techniques. In these studies, simple visual machines (either real robots or simulated agents existing within virtual realities) have been evolved to perform a variety of behaviours mediated by vision or other distal sensing (e.g. sonar, infrared (IR) proximity detectors). Typically, the sensorimotor 'controllers' of these machines are parallel distributed processing systems: commonly, artificial neural networks simulated on a fast serial computer, but also in at least one case (Thompson 1995) real parallel asynchronous analogue electronic circuits. In these studies there is *no precommitment* to any particular representational scheme: the desired behaviour is specified, but there is minimal specification of the mechanism required to generate that behaviour. In the

following three sections we give (i) a brief introduction to artificial evolution, (ii) some examples of artificially evolved simple visual machines, and (iii) then discuss further the issue of representation in these systems.

(a) *Artificial evolution*

Artificial evolution encompasses a number of computational optimization or satisficing techniques which draw inspiration from biological evolution. Only the simplest form of 'genetic algorithm' will be explained here, with specific reference to developing sensorimotor controllers for simple visual machines; for further details, see, for example Goldberg (1989).

In order to apply a genetic algorithm it is necessary to first formulate an *encoding scheme* and a *fitness function*. The encoding scheme is a method of encoding the designs of sensorimotor 'controller' mechanisms (and possibly also the sensor and motor morphology) as strings of characters from a finite alphabet, referred to as 'genomes'. The fitness function takes the spatiotemporal pattern of behaviour of a given individual controller (decoded from a given genome) over one or more trials, and assigns that individual a scalar value which is referred to as its *fitness*, such that desirable behaviours are awarded higher fitness than less desirable behaviours.

The system is initialized by creating a 'population' of individuals, each with a randomly generated genome. The system then enters a loop: all individuals are tested and assigned a fitness score. Individuals with higher fitness values have a greater chance of being selected for *breeding*. In breeding, the genomes of two parents are mixed in a similar manner to recombinant DNA transfer in sexual reproduction, and extra variation is introduced by 'mutations' where characters at randomly-chosen positions on the genotype are randomly 'flipped' to some other character from the genome-alphabet. Sufficiently many new individuals are bred to replace the old population, which is then discarded. Following this, the new population is tested to assign a fitness to each individual. In each cycle of testing the population and breeding a replacement is referred to as one generation, and generally a genetic algorithm runs for a pre-set number of generations, or until the best or average fitness in the population reaches a plateau.

If parameters such as the mutation rate, fitness function, and selection pressure are all set correctly, then typically fitness increases over a number of generations: at the end of the experiment, the best individual genome encodes for a useful design. The final evolved design can then be implemented and analysed to determine how it functions.

In evolving sensorimotor controllers, a variety of possible 'building blocks' can be employed: for a comprehensive review and critique, see Matarić & Cliff (1995). In many of the systems discussed in the next section, continuous-time recurrent neural networks (CTRNNs) are employed: these are artificial neural networks composed of 'neurone' units with specified time-constants giving each neurone an intrinsic dynamics. The primary reasons for employing

such neural networks are (i) their sigmoidal activation function allows them to approximate a very wide class of mathematical functions; (ii) their recurrent connections allow them to maintain their internal state; and (iii) there is a theoretical result which suggests that, appropriately configured, they can approximate a very large class of continuous dynamical systems with arbitrary accuracy. (See Beer (1995*b*) for further details.)

The evolved simple visual machines described below are all both *embodied* and *situated* within an environment: the emphasis is on the evolution of entire sensory-motor coordination mechanisms or processing pathways, constrained only in terms of the fitness of the observable behaviour of the agent. This contrasts with many artificial neural network models, where the constraint is that (either by learning or evolution) the network is capable of making appropriate mappings from a given input representation to a given output representation: modelling entire sensorimotor pathways has a significant impact on the semantics of any representations within the system, see Cliff (1991, 1995).

(b) Examples

As far as we are aware, the first case of an evolved artificial agent using distal sensing was the simulation study by Cliff *et al.* (1993*a*) (see also Cliff *et al.* 1993*b*). In this work, CTRNNs were evolved, along with the specification of the angle of acceptance and physical arrangement of the visual sensors on the robot body. Only two simulated photodetectors (i.e. two 'pixels') were used, but the robot was successfully evolved to visually navigate its way to the centre of a simple arena: a closed circular room with a white floor and ceiling, and a black wall.

Subsequently, Harvey *et al.* (1994) evolved CTRNNs for real-time control of a robot camera head moving in another visually simple environment. The head was mounted with touch sensors and a low-bandwidth charge-coupled device video camera. Networks with three circular receptive fields sampling the input video stream were evolved, with the position and radius of the receptive fields under genetic control. The networks were selected on the basis of their ability to approach a triangular visual target, and avoid a rectangular target: a simple visual categorization task.

Floreano & Mondada (1994) evolved feed-forward neural networks for a simple robot with an eight-pixel input 'image' formed by the inputs of photodetector cells placed around the perimeter of its body (an upright cylinder of height 4 cm and radius 3 cm). These network controllers were evolved to guide the robot through a maze-like environment, attempting to maximize the distance travelled without colliding with the walls of the maze.

Thompson (1995) developed a genetic encoding for electronic circuits composed of digital logic gates, which were asynchronous and recurrently connected, so that the analogue properties of the circuits could be exploited by evolution. The distal sensors were ultrasonic sonars, rather than visual; economical circuits were evolved to allow the robot to guide itself to the centre of a rectangular enclosure using sonar responses.

Jakobi (1994) and Jakobi *et al.* (1995) reported the development of a simulator for the same type of eight-pixel robot used by Floreana & Mondada. They evolved CTRNNs in simulation which could then be successfully transferred to the real robot, generating behaviours which guided the robot towards a light source, while avoiding collisions with obstacles (a task similar to that studied by Franceschini *et al.* (1992)).

Cliff & Miller (1996) evolved CTRNNs for simulated 2D agents using projective geometry to give a 'flatland vision' approximation to visual sensing, with up to 14 pixels in the sensory input vector. Separate populations of 'predator' and 'prey' agents were evolved. The predators were selected for on the basis of their ability to approach, chase, or capture individuals from the prey population; and prey individuals were selected for their ability to avoid being captured by the co-evolving predators.

Finally, Beer (1996) evolved CTRNNs for simulated agents with distal sensing using either five or seven directional proximity detectors: the agents had to perform what Beer refers to as 'minimally cognitive tasks', i.e. behaviours that would usually be assumed to require some form of internal representation or categorization, such as orienting to objects of one particular shape, distinguishing between different shapes, and pointing a 'hand' at certain shapes.

(c) The search for internal representations

All of the evolved simple visual machines discussed above perform tasks that are trivial by the standards of most machine vision research. There is little or no doubt that these tasks could all be solved using a knowledge-based approach, involving a sequence of transformations on appropriate internal representations. Yet the significance of these machines is not the complexity of the problems they solve or the behaviours they exhibit, but rather the way in which their design was produced. In contrast to traditional engineering design techniques, the use of an evolutionary approach with minimal pre-commitments concerning internal architecture or representations makes the question 'What types of representation do these machines use?' an empirical one. That is, we must examine or analyse the evolved designs, generate hypotheses about the representations employed, and test those hypotheses in an appropriate manner. Possibly, the evolutionary process will have resulted in a knowledge-based or model-based solution, in which case appropriate representations will be found; or possibly not.

And it is on this issue that the true significance of these simple visual machines is revealed: as far as we are aware, no analysis of the evolved systems described above has identified the use of representations or knowledge in the conventional (physical symbol system) sense. That is, none of these systems operate by forming a representation of the external environment, and then reasoning with or acting upon that representation (e.g. by comparison with, or reference to, in-built or acquired representations). This is in spite of the fact that a machine-vision engineer, conversant in the methods of knowledge-based vision, could (trivially)

develop an appropriate computational theory for any of these tasks, identify appropriate representations and transformation algorithms to act on them, and specify an implementation in some physical hardware. Evolution, working with primitive building blocks to construct parallel distributed processing architectures for these tasks, just does not do it the knowledge-based way.

This is not to say that the operation of these systems is a mystery. Full causal mechanistic explanations of the evolved systems can be offered via analysis, typically using the tools and language of dynamical systems theory. (For further discussion of the rationale for and use of dynamical systems theory as an alternative to computational/representational accounts of cognition, see Smithers (1992, 1995), Thelen & Smith (1994), Port & van Gelder (1995) and Beer (1995a).) Causal mechanistic explanations are also the ultimate aim of much work in analysing evolved biological systems (Horridge 1977).

For example, the two-pixel controllers evolved to guide a simulated robot to the centre of a circular room (Cliff *et al.* 1993), have been analysed both qualitatively (Cliff *et al.* 1997) and quantitatively (Husbands *et al.* 1995). The behaviour of the robots can be explained and predicted by reference to the dynamics of the agent–environment interaction. The CTRRNs can maintain their internal state, and the state-space of the networks has certain identifiable attractors which correspond to (or are correlated with) certain situations or relationships between the agent and the environment, such as the robot being at the centre of the room. There is a closed sensory-motor loop, in the sense that the changing state of the network is affected by the current and past inputs to the sensors, which are determined by the path the robot takes through the environment, which is in turn determined by the changing state of the network. When the robot is released into the environment at a particular orientation and location, the sensors receive certain light values, which can perturb the state-space trajectory of the CTRNN, which affects the motor outputs, possibly moving the robot, and hence altering the light values subsequently sampled by the sensors. As this state-space trajectory unfolds, the robot can be observed to be moving toward the centre of the circular room, and staying there once it arrives, but there is nothing within the CTRNN that can usefully be described as a representation. There is nothing, for example, corresponding to a stored version of a 'goal state' such as the sensory inputs received when at the centre of the room, or a method for determining, on the basis of comparison with stored values, whether the robot should turn left or right, move forward or reverse, or stop.

Of course, it is famously difficult to prove a negative, and it is beyond the scope of this paper to give a full illustrative example analysis of one of the evolved systems listed above, but a simple thought experiment, adapted from Braitenberg (1984), will serve as a useful illustration. Consider the design for a simple visually-guided wheeled robot with a body plan symmetric about its longitudinal axis. At the front, on the long axis, is a single castor-wheel. At the rear left and rear

right, there are identically sized wheels, attached to independent electrical motors with colinear axles. The robots are differential-steer devices (by altering the angular velocities of the two rear wheels, the robots can travel in arcs of varying radii, either clockwise or anticlockwise). At the front-left and front-right of the robot there is a forward-pointing light sensor. A wire leads from each sensor into a black box where some control circuitry and batteries are hidden. Wires lead from the black box to the two drive motors. Two such robots, marked A and B, are placed in a dark room with no obstacles except for a floor-mounted light-bulb. When the light-bulb is switched on, robot A (which was initially not pointing toward the light-bulb) turns to face the bulb and accelerates toward it, only stopping when it hits it. Meanwhile, robot B (which was initially facing the light-bulb) turns away from the bulb, moving fast at first but then more slowly until it comes gently to a halt. If we were now to ask a knowledge-based vision engineer to theorize about what might be hidden inside the black boxes of robots A and B, s/he would, presumably, in following Marr's three levels of analysis, first formulate a computational theory for each robot, characterizing the performance of each as a mapping from one kind of information to another, and thereby establishing a link from visual information received at the sensors to information concerning appropriate motor outputs. The engineer would then determine the representations for input and outputs, and any intermediate representations, and the algorithm(s) for transforming between them; finally s/he would address issues of how the representations and algorithms can be realized physically. Quite probably, the solution will involve measuring the signals received from the left and right sensors, comparing them (or their difference) to some reference values, and issuing appropriate motor commands on the outcome of the comparison. Given enough time and money, we have no doubt that such controllers could be built and would operate successfully. But, upon opening the black-box controllers on A and B, there is a surprise lurking. The black box in A simply has a wire connecting the left-hand sensor to the right-hand motor, via an appropriate amplifier, and a wire connecting the right-hand sensor to the left-hand motor, again via an amplifier. Similarly, the black box in B has nothing but an amplifier sitting between a wire joining the left sensor to the left motor, and another amplifier between the right sensor and the right motor. All the amplifiers do is ensure that the signals coming from the light sensors are magnified sufficiently to drive the motors: they provide a constant of proportionality, but essentially each motor is driven by a direct connection from one sensor. (Readers familiar with Braitenberg (1984) will recognize A as the contralaterally connected Vehicle 3a, and B as the ipsilaterally connected Vehicle 3b.) This is all it takes to generate the observed behaviours. And the key issue here is that, despite the knowledge-based vision engineer being able to specify representation-manipulating controllers, the actual controllers for these two vehicle robots *use no representations*. Their observable behaviour is a result of the dynamics of

interaction between the agent (robot) and the environment (floor of a dark room and a light-bulb). A complete account of the behaviour of either agent requires treating the agent and environment as coupled (through a sensory motor loop); and there is no useful definition of 'representation' that allows any variable with these coupled systems to be described as a representation *to the agent* of any external object, situation, or event. Of course, this argument rests on the definition of 'representation', a point we return to below.

Just as the Braitenberg vehicles use no representations, so we argue that the artificially evolved simple visual machines discussed previously use no representations. Now it should be noted that, in the majority of cases, the researchers responsible for the evolved simple visual machines are highly doubtful as to whether traditional notions of representation serve any useful purpose in explaining cognitive systems (artificial or natural). Their work is part of a wider movement within the Adaptive Behaviour research community that questions or rejects traditional symbolic notions of representation (for overviews, see for example, Brooks 1991*a,b*).

For this reason, it is pertinent to ask whether (and with all due respect to the researchers involved) representation has not been identified in these machines, because the researchers had a vested interest in not finding any. Put another way: if evolution did produce a design that used internal representations, how would we recognize it? This requires a firm definition of representation: preferably an operational definition (i.e. the specification of a procedure by which an independent third party could establish whether representations are being used or not). It is this issue of attempting to usefully define 'representation' that we turn to in the next section: analysis may identify causal interactions, or high order correlations, but surely a representation is more than just an interaction or a correlation?

4. WHAT IS IT LIKE TO BE A REPRESENTATION?

Harvey (1992, 1996) argues that the only meaningful sense in which internal representations can be discussed in cognitive systems is to recognize that the verb 'represent' should be treated as a four-place predicate: that P is used by Q to represent R to S. For example, the character string 'mast' is used by writers of English to represent 'long upright pole on which the sails of a ship are carried' to English readers. But people writing in Serbo-Croat use exactly the same character string to represent 'ointment, fat, or lard' to Serbo-Croat readers: as Harvey emphasizes, Q and S are necessary to allow for the same P representing different Rs to different P-using communities. So, to talk of representations in vision (and anywhere else), we need to determine who or what are filling the roles of Q and S. We, as external observers, can safely talk of patterns of activity in the nervous system as representing external objects or events to us: Q and S are us humans engaging in a discourse where it is socially agreed that the neural

activity patterns (P) represent some external object(s) or event(s) (R). But to talk about the patterns being representations used by the agent (robot or animal) implies that an agent-within-the-agent is somehow 'reading' these representations: if P is some representational pattern of activity on a defined set of neurones, and we say that P represents some external object or event R, then we should also be able to specify Q and S. If we want to define Q as the collection of neurones over which the pattern P is detected, then what is S? Some other part of the agent's neural system, excluding the neurones in Q? If that is the case, then it is not the agent as a whole that is using the representations: the agent becomes decomposed into a community of sub-agents, forming, using, and exchanging representations. Of course, systems designed by traditional engineering techniques *can* be described this way. But applying this style of description to an *evolved* agent requires care: Harvey's reasoning implies that, unless used carefully, explanation of an agent's neural mechanisms in terms of representations used 'by the agent' can hide an implicit homunculus: the (sub-)agent that reads the representation. And with this homunculus comes the manifest danger of infinite regress.

One means by which a representation can be distinguished from a correlation is by noting that Harvey's argument implies that representations are essentially linguistic (i.e. they form an *interlingua* between representation-using agents or entities). A representation should therefore be normative: it should at least offer the opportunity to misrepresent; to more or less correctly capture some external state of affairs. In the simple visual machines discussed above, there is no representation because there is no possibility of misrepresentation. We, the external observers, can point to the activity patterns and refer to them as representations in explaining the system, and be right or wrong to varying degrees about what those patterns represent. But to talk of the agent *using* the representations is to confuse patterns of activity which represent something else, and patterns of activity which actually constitute the agent's perceptual or experiential world, a point forcefully made by Brooks & Stein:

'There is an argument that certain components of stimulus-response systems are "symbolic". For example, if a particular neuron fires—or a particular wire carries a positive voltage—whenever something red is visible, that neuron—or wire—may be said to "represent" the presence of something red. While this argument may be perfectly reasonable as an observer's explanation of the system, it should not be mistaken for an explanation of what the agent in question believes. In particular, the positive voltage on the wire does not *represent* the presence of red *to the agent*; the positive voltage *is* the presence of something red as far as the robot is concerned.' (Brooks & Stein (1994), original emphasis.)

It could be argued that the simple systems studied so far are merely demonstrations that 'knowledge' and 'structured representations' are not required for such simple tasks, but will be necessary for more complex tasks. We disagree.

Rather, we argue that 'knowledge' and its 'representation' may be nothing more than constructs from folk-

psychology. We maintain that these terms are best viewed as place-holders for yet-to-be-identified causal mechanistic interactions: philosophically, this is a position of eliminative materialism such as that first proposed by Churchland (1979, 1989) and subsequently argued for by Smithers (1992). Such a position also has clear parallels with the work of Braitenberg (1984), who demonstrated that mentalistic notions such as 'fear' and 'aggression' are easily imputed by external observers of his Vehicle series of simple visual machines, two of which were introduced in the thought experiment discussed earlier. Briefly, Braitenberg's argument is that human observers ascribe mental states to the vehicles when describing their actions (e.g. 'robot A approaches light-bulbs *aggressively*' or 'robot B is *frightened* of light and turns away from it'), yet these mentalistic terms have no place in explanations of the causal mechanisms involved in the generation of those behaviours.

In sum, our position is that 'knowledge' and its 'representation' are useful notions at levels of explanation higher than the causally mechanistic; and in particular are valuable when analysis has yet to uncover the causal mechanisms involved in the visual processing mechanisms of interest. But that when an evolved system is fully analysed at the causal mechanistic level, there is no useful place for these terms.

For this reason, we find it hard to agree with statements such as that from the synopsis of this Discussion Meeting: '. . . visual systems acquire and use knowledge in many ways. It is encoded into . . . visual systems by evolution and perhaps still more by individual experience.' For evolved simple visual machines, although it is useful for us to talk of 'knowledge encoded into visual systems' *before* we analyse them, once the analysis is complete and we have a causal mechanistic explanation of the system, there are *only* the interaction dynamics: there is nothing we can point to (or wave our hands over, for fans of 'distributed representations') as the knowledge in the system. It is as elusive as the Ghost In The Machine.

5. IS THE SAME TRUE OF ANIMALS?

Given the existence of evolved artificial systems which exhibit visually guided behaviours yet employ no representations, it is compelling to consider whether similar systems exist in the natural world. Although there is no animal for which a complete analysis (comparable to the analyses of the artificial systems enumerated above) is available, we discuss below some suggestive results from phylogenetically diverse animals.

The visual systems of insects, especially the dipteran flies, have been subjected to extensive studies. Examples include fruit-flies such as *Drosophila melanogaster* (e.g. Wolf & Heisenberg 1991), hover-flies such as *Syritta pipiens* (e.g. Collett & Land 1975*a*), and house-flies such as *Musca domestica* (e.g. Reichardt & Guo 1986) or *Fannia canicularis* (e.g. Land & Collett 1974).

These are, probably, the natural systems for which it is most realistic to attempt a complete causal mechanistic explanation of the couplings between (visual)

sensors and motors. Hence, if vision by definition involves the formation and manipulation of representations, these are also the animals in which we are most likely to be able to identify the neural realization of those representations.

From the reflex loops governing take-off and landing responses or optomotor flight stabilization, through the servo systems underlying the chasing or tracking of one fly by another, to the use of visual landmarks for navigation, there exist published accounts of information-processing or control-theoretic analyses, extensive behavioural studies, and relatively rich neurological data from identifiable individual visual interneurones.

Yet to cast these analyses within a 'model-based' or 'knowledge-based' framework would be, surely, to reduce the notions of 'model' or 'knowledge' to vacuity. Consider conspecific-chasing behaviours: for a full causal mechanistic analysis, it is necessary to acknowledge that much of the 'knowledge' about chasing flies of the same species is 'represented' in *the entire design of the animal*. From the anatomy and optics of the eye, through the neural dynamics of the relevant sensorimotor pathways, to the kinematics of the flight motor system, and indeed the aerodynamics of the whole fly: a full account will treat the fly as a subsystem within the coupled dynamical system formed by the interaction of the agent and its environment. (Here the agent is the chasing fly, and the environment is everything else, the space through which the fly is chasing its target, and any relevant objects in that space; the most relevant of which is the target object, which will usually be a conspecific fly but might be many other things, such as flies of another species, distant birds, or peas thrown by nearby biologists (e.g. Collett & Land 1975*b*).) Presumably the 'knowledge' of important system parameters (e.g. the fly's body shape, its moments of inertia and coefficients of friction for both angular and linear acceleration, etc.) is somehow 'represented' in the neural processes responsible for sensory-motor coordination. But such loosely-sketched representations often prove elusive when we consider *how* the representations might be identified within the system.

Again, we do not deny that external observers can derive elegant and useful computational-level analyses of the task faced by the chasing fly, and that these analyses may involve variables which represent to us (the observers) cogent factors in the environment. This is our privilege as external observers. The fly, unable to adopt the perspective of an external observer, has no access to the representations or knowledge that we humans might invoke when explaining the fly-chasing system to other humans. To talk of knowledge or representations being encoded or compiled by evolution into the body design of the fly is to homuncularize either the fly, the evolutionary process, or both.

To reiterate our argument: a priori, one could construct a knowledge-based vision system which delivers representations appropriate to the control of chasing behaviour, but instead it appears that real flies are a collection of neat tricks that exploit the simplicities and regularities of the environment and the required behaviour, thereby circumventing the need

for a full representation-manipulating vision system. That is, it appears that flies do not *actually* use representations, even though they *could*. A possible rejoinder to this is to agree that flies use no representations, but to argue that more complex animals will *have* to form and manipulate representations in virtue of the complexities of either their environments, the behaviours required of them, or both. We have some sympathy for this position (because it admits that there are no representations in flies), but there are studies of animals more complex than flies which, again, we take as an indication that structured knowledge-based representations may not be involved: we briefly review some of these below.

Consider the numerous studies of so-called 'time-to-contact' behaviours, where the time remaining before impact of a seeing animal with some object or surface plays an important role in exhibiting a desired behaviour (often because the behaviour has to be executed or initiated some time *before* the moment of contact). A clear example is Lee & Reddish's (1981) study of wing-folding in the gannet, *Sula bassana*: hunting gannets dive into the sea, from considerable cruising altitudes, to catch fish. The gannet's speed when it hits the water (at near-vertical angles) can be as high as 24 m s^{-1}. To avoid injury, the gannet folds its wings into its body before impact with the sea surface. But when the wings are folded the gannet has greatly reduced aerodynamic control: it is essentially ballistic and hence it cannot make any final adjustments to its flight path, and so is unable to compensate for any last-moment evasive moves by the fish. In simple but extreme terms, if it folds too late, the gannet breaks its wings, and if it folds too early, the gannet goes hungry. Clearly, the ability to accurately judge the time-to-contact with the sea-surface allows the gannet to commence folding at a time t_{fold} seconds before impact, where t_{fold} is also the time taken to fold the wings from a steering position to a safe streamlined pose.

Now it is certainly not impossible that the gannet's nervous system is forming and manipulating appropriately structured internal representations of the external 3D environment, as would be required of a model-based account. But there is a persuasive argument that this is not the case: Lee (1980a) argued that a parameter τ, being the quotient of the rate of expansion of a point on the retinal image and the distance of that point from the pole of the optic flow-field, gives an accurate measure of time-to-contact of the surface. The τ measure is particularly easy to derive if there is a log-polar sampling of the retinal image (e.g. Wilson 1983). Thus, although time-to-contact *could* be derived using a knowledge-based approach, the available evidence is best accounted for by reference to a simple metric, realizable in image-space (i.e. by a succession of retinotopically projected neural sheets), being employed.

Now, once again, defenders of the knowledge-based or representational viewpoint may want to argue that wing-folding is sufficiently important to the survival of gannets that evolution has 'encoded' the relevant knowledge and representations into the gannet visual system. Presumably the 'knowledge' concerns the utility of τ as an indicator of time-to-contact, and the

ease with which it can be derived from an appropriately sampled optic flow-field. But, in the absence of clear definitions of Harvey's Q and S for the diving gannet, to talk of representations within the system is to homuncularize either the gannet or the evolutionary process. Alternatively, it might be conceded that the exploitation of regularities in the gannet's visual environment (i.e. the numerator and denominator in τ) does not constitute a representation-using system, and we need to look at more complex animals or agent–environment interactions. Yet there is a growing body of comparable data from studies of human subjects engaging in a variety of visually mediated behaviours which are acquired and of little evolutionary significance (in the sense that the behaviours are unlikely to have played a part in selection pressures that shaped the human visual sensorimotor system). In tasks such as catching tennis balls (Lee 1980b), striking the take-off board on a long-jump track (Lee et al. 1982), braking or steering automobiles (Lee & Lishman 1977), and leaping up to punch falling volley-balls (Lee et al. 1983), there is evidence that the use of simple features or metrics of the flow field, including τ, can account for the fast reaction times involved, in a far more parsimonious manner than any account involving the formation and manipulation of structured representations. The similarities between these results and Gibson's (1979) influential arguments for 'direct perception' are manifest.

Even in cases where the reaction times are not an issue, manipulation of monolithic structured representations is questionable in several cases where sufficient data is available to form the basis for alternative accounts. We briefly summarize here two exemplar bodies of work: computational neuroethological studies of visually mediated behaviours in frogs and toads, and recent machine vision work on using high order statistical correlations in image space for a variety of tasks.

The first involves an ongoing series of experiments using computer simulations, behavioural studies, and invasive neuroscience in which a team led by Michael Arbib have developed sophisticated computational models of the neural visuomotor mechanisms underlying predation in frogs and toads (e.g. Arbib 1987; Corbacho & Arbib 1995; Cervantes-Pérez 1995). In brief, behavioural studies (e.g. Lock & Collett 1979) have explored the responses of these animals when faced with the task of moving to within snapping distance of an initially distant food item (the 'prey'), given the presence of a 'barrier'; often either a paling fence or a wide, deep chasm. Computational models, drawing heavily on the available neuroscience data (e.g. Ewert 1987), are used to generate action sequences for 'virtual frogs' situated within simulated prey–barrier environments. The behaviour of the virtual frogs can then be compared to the real animals, thereby suggesting additional refinements to the model or further neuroscience experiments. For the purposes of this discussion, the key indication from this body of work is that separate neural pathways are maintained for processing 'prey' and 'barrier' information, and that any conflicts between the desire to approach prey

and the need to avoid the barrier are resolved very late in the neural pathway, close to the initiation of motor schemas. This is in marked contrast with what would be expected from a knowledge-based approach: presumably this would require the frog to form an internal representation of the external environment, including the prey, the barrier(s), and possibly also the frog itself; some reasoning or planning mechanism(s) would then manipulate this representation to determine one or more possible paths to the prey, one of which would be selected for execution. Once again, this is an internally consistent way of doing things and, in principle, a machine could be constructed along these lines. But, unfortunately, all the available evidence indicates that frog and toad visual systems are not built the knowledge-based way.

The second example comes from machine vision studies where tasks that might otherwise be achieved using 3D model-based techniques, including the representation of 3D shape and volume, are solved using approaches which employ multivariate statistics *in the 2D space of the image*. In summary, these methods involve applying statistical techniques such as principal components analysis (PCA) to vectors of points systematically taken from significant contours in the image. The statistical techniques give the primary modes of variation of these contour-points in image space and, crucially, these primary modes of variation are often in close correspondence with variations in the 2D projection of a 3D object as the pose of the object relative to the viewer is altered. That is, the 2D image statistics capture regularities in the projected images of 3D objects in such a way that, to a fair approximation, the 2D statistical model can be used to perform tasks that might otherwise be assumed a priori to require internal representations of 3D shape, volume, etc. Examples of work in this area include Baumberg (1995), Baumberg & Hogg (1996), Lanitis *et al.* (1995) Again, this is not to say that a 3D model-based approach would not be able to perform the task: the work of Baumberg (1995), using image-space statistical techniques to track movie sequences of walking people, complements Hogg's (1983) earlier work on using knowledge-based vision to perform much the same task. But, given the ease with which artificial neural networks can approximate multivariate statistical techniques such as PCA, it is tempting to ask whether real neural networks perhaps employ high order correlations in 2D image space to circumvent the complexity of manipulating internal representations of 3D objects. We see this as a provocative question which can only be addressed by further research, but statistical arguments have been presented as powerful alternatives to representational accounts of lower order visual processes (e.g. Srinivasan *et al.* 1982).

The examples we have given here, from studies of insects, amphibians, birds, and humans, are by no means conclusive proof of our arguments. However, we believe that they are significant and persuasive because, although all of the visually mediated tasks involved could be performed using a knowledge-based approach, the available evidence indicates that they are not. In situations where an a priori consideration of the

task from a knowledge-based vision perspective might lead an external observer or designer to posit the need for structured internal representations, reconstructing the external world, the best a posteriori explanation may be significantly different, employing either no representations, or representations very different from those assumed to be useful on the basis of successful engineering practices in machine vision.

6. SUMMARY: VISION WITHOUT KNOWLEDGE?

It is easy to conjecture the need for knowledge and its representation in vision either when introspecting, as is seen in the experiments of Shepard & Metzler; or when applying divide-and-conquer approaches to the problem of designing a computational vision system, as witnessed in the Marr pipeline; or when dealing with the incomplete data offered by neuroscience, as happens when Marr's three-level methodology is applied to analysing animal vision systems.

But preliminary experience with analysing evolved artificial visual systems indicates that, possibly, the utility of descriptions and explanations involving knowledge and its representation recede as analysis progresses. A priori hypotheses involving the representation and manipulation or mobilization of knowledge are undoubtedly useful for motivating discussion and experimentation, but as more is made known about the mechanisms involved, so the places where the knowledge might be represented or encoded-in recede, and when the analysis is complete, knowledge and its representation are hard to identify in meaningful terms, just as 'aggression' and 'fear' play no part in explaining Braitenberg's vehicles once the lid of the black box is opened.

Our intention in this paper has simply been to highlight the problems that arise when the language of knowledge-based vision is applied to the analysis of evolved machines, either animals or artificial agents. In these systems, where there has been no pre-commitment to any representational scheme, the presence or absence of knowledge and its representation become empirical issues. To pursue the matter further requires at least a consensus on what is meant by 'knowledge' and 'representation'; and better still an operational definition of representation, such that replicable and hence falsifiable experiments can be proposed and conducted.

It is certainly difficult to define the notions of knowledge and its representation sufficiently accurately to provide these operational definitions. But until such operational definitions are agreed upon, arguments that the structured representation of knowledge plays no part in evolved visual systems are unsound. Yet, surely, by the same reasoning, arguments that the structured representation of knowledge *does* play a part in evolved visual systems are *also* unsound. We might be happy to agree that representations have a part to play in explaining vision in animals and other evolved machines, if only we could agree on what a representation is, and on who or what is using those representations.

REFERENCES

Arbib, M. A. 1987 Levels of modelling of mechanisms of visually guided behaviour. *Behav. Brain Sci.* **10**, 407–465.

Ballard, D. H. 1991 Animate vision. *Artif. Intell.* **48**, 57–86.

Baumberg, A. M. 1995 Learning deformable models for tracking human motion. Ph.D. thesis, School of Computer Studies, University of Leeds.

Baumberg, A. M. & Hogg, D. C. 1996 Generating spatiotemporal models from examples. *Image Vision Computing* **14**(8), 525–532. (Also available as University of Leeds School of Computer Studies Research Report No. 95.9.)

Beer, R. D. 1995a A dynamical systems perspective on agent–environment interaction. *Artif. Intell.* **72**, 173–215.

Beer, R. D. 1995b On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behav.* **3**(4), 471–511.

Beer, R. D. 1996 Toward the evolution of dynamical neural networks for minimally cognitive behavior. In *From animals to animats. 4. Proc. 4th int. conf. on simulation of adaptive behaviour* (ed. P. Maes, M. J. Matarić, J.-A. Meyer, J. Pollack & S. W. Wilson), pp. 421–429. Cambridge, MA: MIT Press Bradford Books.

Braitenberg, V. 1984. *Vehicles: experiments in synthetic psychology.* Cambridge, MA: MIT Press Bradford Books.

Brooks, R. A. 1991a Intelligence without reason. In *Proc. 12th int. joint conf. on artificial intelligence (IJCAI-91)*, pp. 139–159. San Mateo, CA: Morgan Kaufmann.

Brooks, R. A. 1991b Intelligence without representation. *Artif. Intell.* **47**, 139–159.

Brooks, R. A. & Stein, L. S. 1994 Building brains for bodies. *Autonomous Robots* **1**, 7–25.

Cervantez-Pérez, F. 1995 Visuomotor coordination in frogs and toads. In *The handbook of brain theory and neural networks* (ed. M. A. Arbib), pp. 1036–1042. Cambridge, MA: MIT Press Bradford Books.

Charniak, E. & McDermott, D. 1985 *Introduction to artificial intelligence.* Reading, MA: Addison-Wesley.

Churchland, P. M. 1979 *Scientific realism and the plasticity of mind.* Cambridge University Press.

Churchland, P. M. 1989 *A neurocomputaional perspective: the nature of mind and the structure of science.* Cambridge, MA: MIT Press Bradford Books.

Cliff, D. 1991 Computational neuroethology: a provisional manifesto. In *From animals to animats. 1. Proc. 1st int. conf. on simulation of adaptive behavior* (SAB90) (ed. J.-A. Meyer & S. W. Wilson), pp. 29–39. Cambridge, MA: MIT Press Bradford Books. (Also available as University of Sussex School of Cognitive and Computing Sciences Technical Report CSRP 162.)

Cliff, D. 1995 Neuroethology, computational. In *The handbook of brain theory and neural networks* (ed. M. A. Arbib), pp. 626–630. Cambridge, MA: MIT Press Bradford Books.

Cliff, D. & Miller, G. F. 1996 Coevolution of pursuit and evasion. II. Simulation methods and results. In *From animals to animats. 4. Proc. 4th int. conf. on simulation of adaptive behavior*, pp. 506–515. Cambridge, MA: MIT Press Bradford Books.

Cliff, D., Harvey, I. & Husbands, P. 1993a Evolving visually guided robots. In *From animals to animats. 2. Proc. 2nd int. conf. on simulation of adaptive behavior (SAB92)* (ed. J.-A. Meyer, H. Roitblat & S. Wilson), pp. 374–383. Cambridge, MA: MIT Press Bradford Books. (Also available as University of Sussex School of Cognitive and Computing Sciences Technical Report CSRP220.)

Cliff, D., Harvey, I. & Husbands, P. 1993b Explorations in evolutionary robotics. *Adaptive Behav.* **2**(1), 71–108.

Cliff, D., Harvey, I. & Husbands, P. 1997 Artificial evolution of visual control systems for robots. In *From living eyes to seeing machines* (ed. M. Srinivisan & S. Venkatesh), pp. 126–157. Oxford University Press.

Cliff, D., Husbands, P., Meyer, J.-A. & Wilson, S. (eds) 1994 *From animals to animats. 3. Proc. 2nd int. conf. on simulation of adaptive behavior (SAB94).* Cambridge, MA: MIT Press Bradford Books.

Clocksin, W. F. & Mellish, C. S. 1984 *Programming in Prolog.* 2nd edn. Berlin: Springer.

Collett, T. S. & Land, M. F. 1975a Visual control of flight behaviour in the hoverfly, *Syritta pipiens* L. *J. Comp. Physiol.* **99**, 166.

Collett, T. S. & Land, M. F. 1975b Visual spatial memory in a hoverfly. *J. Comp. Physiol.* **100**, 59–84.

Corbacho, F. J. & Arbib, M. A. 1995 Learning to detour. *Adaptive Behav.* **3**(4), 419–468.

Dreyfus, H. L. 1979 *What computers can't do,* 2nd edn. New York: Harper & Row.

Dreyfus, H. L. 1981 From micro-worlds to knowledge representation: AI at an impasse. In *Mind design: philosophy, psychology, artificial intelligence* (ed. J. Haugeland), pp. 161–204. Cambridge, MA: MIT Press Bradford Books.

Ewert, J.-P. 1987 Neuroethology of releasing mechanisms: prey-catching in toads. *Behav. Brain Sci.* **10**, 337–405.

Floreano, D. & Mondala, F. 1994 Automatic creation of an autonomous agent: genetic evolution of a neural-driven robot. In *From animals to animats. 3. Proc. 3rd int. conf. on simulation of adaptive behavior (SAB94)*, pp. 421–430. Cambridge, MA: MIT Press Bradford Books.

Franceschini, N., Pichon, J.-M. & Blanes, C. 1992 From insect vision to robot vision. *Phil. Trans. R. Soc. Lond.* B **337**, 283–294.

Gibson, J. J. 1979 *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Goldberg, D. E. 1989 *Genetic algorithms in search, optimization, and machine learning.* Reading, MA: Addison-Wesley.

Gonzalez, A. J. & Dankel, D. D. 1993 *The engineering of knowledge-based systems.* Englewood Cliffs, NJ: Prentice-Hall International.

Harnad, S. 1990 The symbol grounding problem. *Physica* D **42**, 335–346.

Harvey, I. 1992 Untimed and misrepresented: connectionism and the computer metaphor. Technical report CSRP 245, University of Sussex School of Cognitive and Computing Sciences.

Harvey, I. 1996 Untimed and misrepresented: connectionism and the computer metaphor. *AISB Quart.* **96**, 20–27.

Harvey, I, Husbands, P. & Cliff, D. 1994 Seeing the light: artificial evolution; real vision. In *From animals to animats. 3. Proc. 3rd int. conf. on simulation of adaptive behavior (SAB94)* (ed. D. Cliff, P. Husbands, J.-A. Meyer & S. W. Wilson), pp. 392–401. Cambridge, MA: MIT Press Bradford Books.

Hogg, D. C. 1983 Model-based vision. A program to see a walking person. *Image Vision Comput.* **1**(1), 5–20.

Horridge, G. A. 1977 Mechanistic teleology and explanation in neuroethology: understanding the origins of behaviour. In *Identified neurons and behaviour of arthropods* (ed. G. Hoyle), pp. 423–438. New York: Plenum Press.

Horswill, I. D. 1993 Specialization of perceptual processes. Ph.D. thesis, AI Lab, MIT.

Husbands, P., Harvey, I. & Cliff, D. 1995 Circle in the round: state space attractors for evolved sighted robots. *Robotics Autonomous Syst.* **15**, 83–106.

Jakobi, N. 1994 Evolving sensorimotor control architectures in simulation for a real robot. M.Sc. thesis, University of Sussex School of Cognitive and Computing Sciences.

Jakobi, N., Husbands, P. & Harvey, I. 1995 Noise and the reality gap: the use of simulation in evolutionary robotics. In *Advances in artificial life. Proc. 3rd European conf. on artificial*

life (ed. F. Morán, A. Moreno, J. J. Merelo & P. Chacón), pp. 704–720. Berlin: Springer.

Land, M. F. & Collett, T. S. 1974 Chasing behaviour of house-flies (*Fannia canicularis*). *J. Comp. Physiol.* **89**, 331–357.

Lanitis, A., Taylor, C. J. & Cootes, T. F. 1995 Automatic identification of human faces using flexible appearance models. *Image Vision Comput.* **13**(5), 393–401.

Lee, D. N. 1980*a* The optic flow field: the foundation of vision. *Phil. Trans. R. Soc. Lond.* B **290**, 169–179.

Lee, D. N. 1980*b* Visuomotor coordination in space-time. In *Tutorials in motor behavior* (ed. G. E. Stelmach & J. Requin). North-Holland.

Lee, D. N. & Lishman, J. R. 1977 Visual control of locomotion. *Scand. J. Psychol.* **18**, 224–230.

Lee, D. N. & Reddish, P. E. 1981 Plummeting gannets: a paradigm of ecological optics. *Nature* **293**, 293–294.

Lee, D. N., Lishman, J. R. & Thompson, J. A. 1982 Regulation of gait in long-jumping. *J. Exptl Psychol: Human Perception Performance* **8**, 448–459.

Lee, D. N., Young, D. S., Reddish, P. E., Lough, S. & Clayton, T. M. 1983 Visual timing in hitting an accelerating ball. *Q. J. Exptl Psychol.* **35**A, 335–346.

Lock, A. & Collett, T. 1979 A toad's devious approach to its prey: a study of some complex uses of depth vision. *J. Comp. Physiol.* **131**, 179–189.

Maes, P, Matarić, M. J., Meyer, J.-A., Pollack, J. & Wilson, S. W. (eds) 1996 *From animals to animats. 4. Proc. 4th int. conf. on simulation of adaptive behavior.* Cambridge, MA: MIT Press Bradford Books.

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

Matarić, M. J. & Cliff, D. 1995 Challenges in evolving controllers for physical robots. *Robotics Autonomous Syst.* **19**(1), 67–83.

Meyer, J.-A. & Wilson, S. W. (eds) 1991 *From animals to animats. 1. Proc. 1st. int. conf. on simulation of adaptive behavior (SAB90).* Cambridge, MA: MIT Press Bradford Books.

Meyer, J.-A., Roitblat, H. & Wilson, S. W. (eds) 1993 *From animals to animats. 2. Proc. 2nd int. conf. on simulation of adaptive behavior (SAB92).* Cambridge, MA: MIT Press Bradford Books.

Nelson, R. C. 1991 Introduction. *Int. J. Computer Vision* **7**(1), 5–9.

Newell, A. & Simon, H. A. 1976 Computer science as empirical enquiry: symbols and search. *Communications Ass. Comput. Machinery* **19**(3), 113–126.

Nilsson, N. J. 1982 *Principles of artificial intelligence.* Berlin: Springer.

Pentland, A. P. 1986 *From pixels to predicates: recent advances in computational and robotic vision.* Norwood, NJ: Ablex Publishing.

Port, R. & van Gelder, T. (eds) 1995 *Mind as motion: explorations in the dynamics of cognition.* Cambridge, MA: MIT Press Bradford Books.

Reichardt, W. E. & Guo, A. 1986 Elementary pattern discrimination (behavioural experiments with the fly *Musca domestica*). *Biological Cybernetics* **53**, 285–306.

Shepard, R. N. & Metzler, J. 1971 Mental rotation of three-dimensional objects. *Science* **171**, 701–703.

Smithers, T. 1992 Taking eliminative materialism seriously: a methodology for autonomous systems research. In *Towards a practice of autonomous systems. Proc. 1st European conf. on artificial life (ECAL91)* (ed. F. J. Varela & P. Bourgine), pp. 31–40. Cambridge, MA: MIT Press Bradford Books.

Srinivasan, M. V., Laughlin, S. B. & Dubs, A. 1982 Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond.* B **216**, 427–459.

Sullivan, G. D. 1992 Visual interpretation of known objects in constrained scenes. *Phil. Trans. R. Soc. Lond.* B **337**, 361–370.

Taylor, A., Gross, A., Hogg, D. C. & Mason, D. C. 1986 Knowledge-based interpretation of remotely sensed images. *Image Vision Comput.* **4**, 6783.

Thelen, E. & Smith, L. 1994 *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: MIT Press Bradford Books.

Thompson, A. 1995 Evolving electronic robot controllers that exploit hardware resources. In *Advances in artificial life. Proc. 3rd European conf. on artificial life* (ed. F. Morán, J. J. Moreno & P. Chacón), pp. 640–656. Berlin: Springer.

Warrington, E. K. & Taylor, A. M. 1973 The contribution of the right parietal lobe to object recognition. *Cortex* **9**, 152–164.

Warrington, E. K. & Taylor, A. M. 1978 Two categorical stages of object recognition. *Perception* **7**, 695–705.

Wilson, S. W. 1983 On the retino-cortical mapping. *Int. J. Man-Machine Studies* **18**, 361–389.

Winston, P. H. & Horn, B. 1980 *LISP.* Reading, MA: Addison-Wesley.

Wolf, R. & Heisenberg, M. 1991 Basic organization of operant behaviour as revealed in *Drosophila* flight orientation. *J. Comp. Physiol.* A **169**, 699–705.