NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

ARTIFICIAL INTELLIGENCE AND ANIMAL PSYCHOLOGY

Margaret. A. Boden

Cognitive Science Research Paper

Serial no: CSRP 016

The University of Sussex
Cognitive Studies Programme
School of Social Sciences
Falmer
Brighton BN1 9QN

ARTIFICIAL INTELLIGENCE AND ANIMAL PSYCHOLOGY

Margaret A. Boden

I. Introduction

Honey-bees returning from new territory are in the happy position of being able to communicate unambiguously the desirability of what they found there. Students of animal behaviour who wander from the familiar paths of ethological and psychological research into the field of artificial intelligence (A.I.) will not find themselves similarly blessed. For there is both good news and bad to be broadcast to their fellows after such a foray.

The good news is that animals must indeed be credited with the ability to form symbolic representations; that this is so even in the absence of communicative behaviour on their part; that there exists a source of concepts for clearly articulating the structure of and transformations between different representations; and that there exists (though at a great distance) the possibility of a principled matching of varying content and function to distinct representational forms. The bad news is that the problems of formulating plausible hypotheses about animals' representations are even more complex than is generally believed; that, given any such hypothesis, the possibility of testing it is more problematic than might at first appear; and that the most puzzling feature (though not all features) of consciousness remains unresolved by this approach as by all others.

In Section II, I say something in general terms about A.I. as the study of representation, and explain why it suggests that we must attribute symbolic representations even to non-communicating animals. Next (in Section III), I relate some problems about motor action and perception in animals to examples of current work in A.I. These problems are typical of those raised within "cognitive ethology," a term recently coined to cover studies of the psychological competence of animals, such as the work on chimps directed by D. Premack or D. M. Rumbaugh, or comparable work on other species. I shall discuss examples concerning both motor action and perception. In Section IV, I outline the reasons for being doubtful about the validatory power of animal experimentation—even work as fascinating as the recent studies of chimps just mentioned. And in Section V, I say a little about the problem of ascribing conscious states to animals, a problem which is addressed by some self-styled cognitive ethologists. I end as I begin, by emphasizing that A.I. is an unripe fruit on the scientific vine, too immature as yet to offer the satisfactions of vintage wine. But where there can be no hope of quenching thirst, there may yet be a chance to wet one's palate.

II; A-2* £S the Study of Representation

Years ago, I saw in the pages of *Punch* a cartoon more memorable than most (I have redrawn it in Figure 1TT It showed

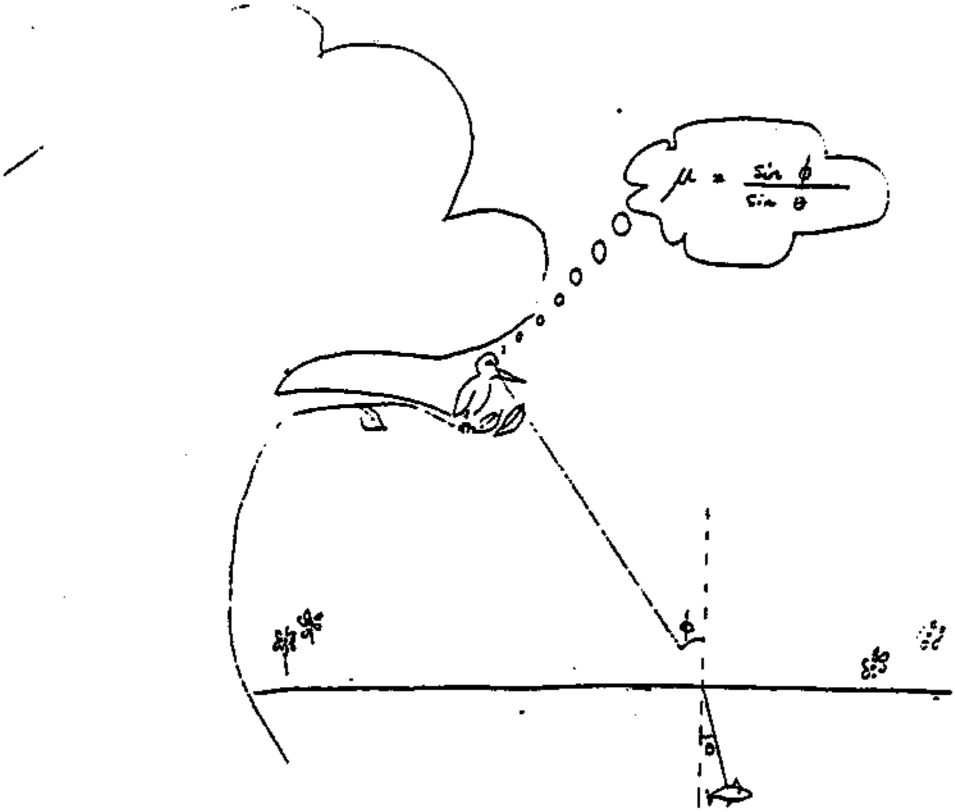

FIGURE i

a kingfisher sitting on a willow-branch, staring at a fish in the river below, and thinking to itself, ". . . . ." This cartoon is no mere triviality, for it is a reminder of some deeply puzzling questions. How does the kingfisher manage to catch the fish, no matter (within limits) where it is in the water? Unlike some birds, it does not dive vertically into the water, nor does it pursue the fish while under water. Kingfishers are plunge-divers, who go rapidly straight to the target. Given that the kingfisher has never heard of Snell's law, does it have to go through some alternative process of computation to adjust its angle-of-dive appropriately—and if so, what? A less obvious puzzle is how the bird manages to identify part of the scene as a fish, or as food, in the first place, and how it is thereupon led to take appropriate action (that is, how does it know that it should dive, irrespective of how steep the dive should be)?

All these puzzles concern the information being used by the animal, and the way in which the animal is using it. One might expect, then, that A.I. should be somehow relevant. A.I. is a recent branch of information-science that is suited to the needs of ethology or theoretical psychology because it defines a wide range of qualitatively distinct and structurally complex symbolic representations and interpretative procedures. The computational concepts used in A.I. are concerned with the reception, storage, transformation, interpretation, and use of information by information-processing systems which employ and construct symbols—and symbol-manipulation procedures--of many distinct kinds.

It is often regarded as problematic whether or not animals have mental representations, or use symbolic systems or languages. Sometimes it is even stated categorically that they do not. Piaget, for example, says that animals (like newborn babies) do not have representations, and Chomsky denies that even chimps enjoy language. To some degree, these disputes turn on terminological differences in the use of terms such as "representation" and "language". Thus Piaget acknowledges that animals (and young babies engaging in reflex action) make use of some degree of "consciousness of meaning," that they construct "significations" if not "representations." Similarly, many people define "language" in such a way that only a system of communication between conspecifics could count as language: and many (Chomsky included) regard abstract features such as syntactic structure or individuating reference to past events as necessary to "language". But even setting aside such terminological differences, it remains true that whether or not any animals employ symbolic representations is widely regarded as doubtful.

The lesson of A.I. is that many animals must have both representations and symbolic language, enabling them to interpret stimulus-information sensibly in widely differing contexts and to take appropriate action accordingly. The more flexible the action, the more complex must be the computational resources for monitoring, planning, and scheduling different types of activity. In particular, when the creature has to take account of a wide range of structural differences and similarities between distinct situations (as opposed to concentrating on only one or a few physical parameters), these structural features can only be represented symbolically—for, by hypothesis, they have no physical features in common. This is true whether or not the animal is able also to communicate with its conspecifics, by warning-cries, mating-calls, and the like. And it is true whether or not the animal is able, like humans, to employ a syntactically structured public language, using units of meaning whose semantic import is determined by social conventions rather than by fixed genetic mechanisms. The point of present importance is that even much noncommunicative behaviour has to be understood in computational terms such that internal symbolic processes must be attributed to the creature. Indeed, the interpretation of audible or visible signs, words, or gestures as communications with a certain meaning presupposes the computational mechanisms involved in sensory perception in general. This is why one A.I.-worker has referred to "the primacy of non-communicative language" [2].

This is not to say that all computations carried out by animals are effected symbolically. For example, hoverflies appear to compute their interception paths with conspecifics according to a simply specifiable rule, one which could plausibly be "hardwired" into the flies' brains [1]. Although this rule could be represented and applied within a symbolic system, it is reasonable to suggest that it has been "learnt" by the evolutionary process and is embodied in the flies' neurophysiology. It is significant, however, that the computations concerned are relatively inflexible: the fly in effect assumes that the size and velocity of the target are always those corresponding to hoverflies, and on this rigid (and fallible) basis the creature determines its angle of turn, when initiating its flight, according to the variable approach angle subtended by the target. Moreover, the fly's path cannot be adjusted in mid-flight, there being no way in which the pursuer can be influenced by feedback from the (perhaps unpredictable) movement of the target animal. This rigid behaviour is fairly common in insects, but the higher animals are capable of considerable flexibility in adjusting their behaviour to widely differing (and continuously changing) circumstances, where the relevant "parameters" are structural features rather than physical ones (such as angle-of-approach).

What is meant by "the computational mechanisms involved in sensory perception in general," and the "flexibility" of action, will become clearer in the next section. To put some flesh and blood onto these skeletal remarks, then, let us turn to see how some problems about action and perception raised within cognitive ethology might fruitfully be approached in A.I. terms.

III: Cognitive Ethology and Computational Concepts

In his seminal paper on "the invisible worlds of animals and men," Jacob von Uexkull showed that the task of a cognitive ethology is to articulate the varied Umwelten of different species [3]. To do this, we need to ask what a given species can perceive, and what it can do accordingly. Those aspects of environment or action which a creature does not have the epistemological resources to represent, cannot form part of its cognitive world. Von Uexkull illustrated these points by his unforgettable pictures of the living-room as seen by fly, dog, or man, and of the fish and the boat as seen by a sea-urchin. But, charming though they are, his pictures do not clearly articulate the similarities and differences between the invisible worlds of these species. Work in A.I. might help us to a richer understanding of such matters, as I shall now try to show by reference to examples concerning the planning of action and the perception of the physical world.

The concept of purposive action has often entered ethological discussion. Purposive action is behaviour controlled by a guiding representation of some desired state, whose overall plan allows for obstacles to be overcome by appropriate variations in the activities selected as means to the end (notice that this is an essentially psychological definition, so that purposive behaviour is not the same as behaviour controlled by feedback of the sort studied in classical cybernetics or control-theory) [4]. To be sure, ethologists in the past have been more concerned to deny the relevance of this concept to animal behaviour than to insist on its applicability. Thus in 1937 Konrad Lorenz criticized anthropomorphic attributions of "instinct", saying that "To assume a 'whole-producing,' directive instinct superior to all part reactions could evidently be justified only if the effects of a regulative factor, exceeding the experimentally demonstrable regulative faculty of the single reactions, could be observed" [5].

Since then, entomologists and ornithologists in particular have often identified independently controlled units of behaviour, which in normal circumstances combine to give the appearance of activity that is planned as a whole and dependent on a recognition of complex means-end relationships. D. S. Lehrman's studies of parental behaviour in ring-doves are an example of this sort of analysis [6]. Various hormonal factors and several "social" releasing stimuli interact, so that the behaviour of birds and squabs is reciprocally determined in an appropriate manner. But if the normal sequence is upset the doves do not engage in variation of means, they do not adapt their behaviour intelligently so as to achieve the desirable (though evidently not desired) end-state of a nestful of happy, healthy chicks. (It does not follow, of course, that ring-doves never engage in behaviour guided by desires, nor that the independent parts of the "parental" sequence are not flexible to some degree according to circumstance: but the guiding goal of rearing healthy chicks cannot be posited as an explanation of this sequence.)

Recently, however, primatologists have begun to ask whether the behaviour of apes, at least, may sometimes be directed by plans or strategies guided by an idea of the goal. And some ethologists raise this question also about non-primate, and even non-mammalian, species such as beavers and bees [7]. But it is generally agreed, even by those willing to consider such questions, that they are very difficult to answer. This difficulty rests partly in the fact that psychology has not provided a theoretical vocabulary for expressing the structure of purposive activity. Indeed, for many years Anglo-Saxon psychologists actively discouraged any such endeavour, because of the anti-mentalistic bias of behaviourism. A.I. may be helpful here, for there are already a large number of A.I. programs concerned with planning, in which are defined procedures of varying complexity for comparing current with desired state and selecting activities accordingly [8]. The computational concepts involved offer the beginning of a theoretical taxonomy of plans. Such a taxonomy could aid the behavioural analysis of those forms of animal activity that are apparently purposive, rather than being simply "automatic" or "mechanical" in nature. In a recent publication, Lorenz has cited examples showing that this is a continuous range rather than a bipolar distinction within animal behaviour [9].Computational considerations could help distinguish the different points in the behavioural spectrum.

Many actions of insects are sequential patterns of invariant order which, once started, are "automatically" executed to the bitter end even in inappropriate circumstances. Sometimes there is a degree of flexibility due to local conditions (such as the configuration of the terrain), but there is no feedback of information capable of altering the overall pattern; at best, it can be interrupted, cut short without the possibility of restarting later at the same point. And some other examples of animal action (such as the parental behaviour of Lehrman's ring-doves) are composed of units which follow "mechanically" in a fixed order provided that at each point the relevant releasing stimulus occurs.

A.I. plans are not like this, although most people unfamiliar with A.I. assume that they are. They are hierarchically organized wholes, variable according to circumstance. Many programs have a "heterarchical" control-structure, in which control is widely distributed throughout the system: the sub-programs on various levels can communicate up and down and sideways, so that decisions can be taken at a local level relatively independently of the overall goal of the system as a whole. This type of control-structure (which is often compared to a human committee of experts) makes it easier to effect subtle variations according to context, so that what at a higher level is clearly "one and the same plan" can be interpreted in importantly different ways on different occasions. Various sorts of monitoring activity are employed to schedule different sorts of action and to make adjustments to ongoing action that is not proceeding (in its relation to the problem-environment) as well as it might be.

For example, some programs monitor and adjust the execution of their plans by reference to their internal representation of the preconditions and consequences of different actions. Thus the mobile robot SHAKEY, while executing a plan for moving blocks from one room to another, asks itself at each step whether the plan as executed so far has produced the expected results (which it may not have done if the environment has changed unexpectedly); what portion of the plan needs to be executed next (which may not be the portion initially foreseen, if the previous question was answered in the negative); and whether this next portion can indeed be executed in the current state of the world (if not, a sub-goal may be set up to realize the necessary preconditions). Other planning programs exist with a richer representational power and so a greater flexibility of action. Some can choose in a principled fashion whether or not to commit themselves to a specific ordering of subgoals ahead of time, and accordingly decide sensibly when the time comes to execute the plan. Some can generate an outline plan that omits all reference to detail, and translate this outline into detailed effective action when necessary. Some can anticipate unwanted side-effects and modify the plan accordingly, so as to avoid them or neutralize thier unwelcome aspects. Some can envisage different alternative strategies for achieving a goal, and use both reasoning and empirical enquiry in choosing between them. Some can recognize a *cul-de-sac* and re-enter a strategy at the precise point where it was previously abandoned, possibly generating a new mini-plan for overcoming the local obstacle which (as it remembers) led to its abandonment in the first place. And some can construct a representation of the goals and plan-following of another program, using it to guide interaction between the two systems C103.

Were one to apply the insights gained in the development of these programs to the experiments on chimps done by D. Premack and G. Woodruff, or by D. M. Rumbaugh *e£ a^./* one would be led to ask a number of questions not mentioned by them. For example, how sensitive are chimps to constraints on the temporal ordering of certain units of behaviour in the context of an overall problem, such that *this* sub-unit has to be performed before *that* one? (They clearly are sensitive to such constraints in some degree, since they will often go to fetch a tool before attempting to do the task for which the tool is required.) Does a chimp have the representational complexity to gather together two or three tools, each of which will be needed in the ensuing task? Or must the chimp think about only one step at a time? If it sees another individual attempting the second step before trying the first (where this ordering is mandatory), can the chimp realize and communicate the information that the required step should be taken instead? If a chimp decides to abandon a task, what features influence its decision? Is it capable of coming back to that task at an appropriate moment, and if so can it remember where it was in the task previously, or must it begin again from scratch? Does a chimp ever engage in activity which looks as though it is a preparation for some later task, either in establishing necessary preconditions or in forestalling unwelcome consequences that would otherwise.ensue on later performance of the task in question? If a chimp is interrupted during its problem-solving by some irrelevant occurrence, does it remember the unfinished task, and does it remember what stage it had reached at the time of interruption? And so on, and so on. We must not merely ask whether chimps generate representations of plans, but must distinguish the computationally different types of plan that they might be using in the control of their behaviour, and that they might be attributing to other individuals (whether chimp or human).

Some A.I. workers would echo Lorenz at this point, objecting that one cannot assume that apparently integrated behaviour is controlled by some integrally organized plan, or that flexible, context-sensitive behaviour is guided by a representation of the desired overall result, They would refer to programs called "production systems," in which control rests in a number of largely independent rules, each of which may be acquired in isolation, and each of which expresses a Condition-Action pair. Each rule tests for a certain Condition (in input or in short-term memory) and then carries out the relevant Action (either producing output or altering the contents of short-term memory). This approach is quite different in spirit from the "planning" approach previously described. It is better able to represent the continual shifting of the focus of attention, and also the interruption of behaviour, whereby appropriate action can be instantly taken on the occurrence of an unexpected event. Yet it can model problem-solving behaviour which one might have thought to be controlled by a plan explicitly representing the structure of the task as a whole. However, this purposive structure has to be implicit in the system if it is to model hierarchically integrated behaviour. So, for example, constraints have to be written into the content of the rules, or the priority and/or temporal ordering of the rules have to be constrained, in ways that decrease the independence of the several rules and so go against the spirit of the approach in its pure form [11].

Production systems can sometimes be matched to detailed behavioural protocols, and studied <u>pari passu</u> with experimental results. For example, a system of rules whose subsets generate different patterns of seriation (staircase-building) can be matched to children's motor and verbal behaviour so as to capture a wide range of detailed observations [12]; and a production system for subtraction can model the many commonly observed errors in subtraction sums that children make [13]. These examples show that even a small number of production rules can give rise to performance that is considerably more varied and flexible than the relentless formula-following common in insects or the successive behaviour-triggering seen in Lehrman's ring-doves. And large production systems, incorporating many hundreds of rules, can generate problem-solving performance comparable to subtle and complex human behaviour.

It may be that much animal behaviour, especially non-mammalian forms, could fruitfully be modelled in these terms. For the Condition may be an external environmental condition (temperature, sunrise, or the presence of a fish or a cat), a state of the animal's internal environment (hormonal concentration), or an inner psychological condition (such as the impulse or desire to catch a fish). And the Action may be motor behaviour (as in diving for the fish or fleeing the cat), or psychological processing (as in activating the desire to catch a fish, or checking to see whether it is on the surface or deep in the water). Ethologists might find it useful to 'try to write production systems modelling behaviour in different species, and to enquire into the acquisition (whether genetic or through learning) of individual rules. Running one's set of rules on a computer enables one to test not only their coherence but their implications, for one can systematically omit or alter individual rules and observe the performance so generated. In this way, one might enrich one's understanding of what Lorenz termed "the experimentally demonstrable regulative faculty of the single reactions."

Coordinated with its active aspect (whether this be regulated by an overall plan or by isolable rules), the Umwelt of any animal has a perceptual aspect. For example, many species are assumed by ethologists to enjoy motion-perception and object-concepts of some sort. Just what sort, however, is usually unclear. Even in the human case, the psychological processes underlying motion-perception are not fully understood. Some recent A.I.-based work done by Shimon Ullman suggests computational questions and hypothetical answers that are relevant not only to human vision (Ullman's prime focus), but to animal vision also [14]. Like the psychologist J. J. Gibson, Ullman attempts to show that many perceptual features can be recognized by relatively low-level psychophysiological mechanisms, whose functioning relies on the information available in the ambient light rather than on high-level concepts or cerebral schemata [15]. But unlike Gibson, who posits a "direct" unanalysable perceptual process of "information pick-up," Ullman views this functioning as a significantly complex process intelligible in computational terms.

Ullman reminds us that if two differing views or input-arrays are successively presented to the visual system, then one of several phenomenologically distinct perceptions may arise. We may see an object (visible in the earlier view) disappearing, and being replaced by another one--as in a game of "peekaboo"; we may see one and the same (rigid) object moving, perhaps involving a change in its appearance due to rotation; we may see one and the same object changing in shape so as to be transformed into something different--as the baby that Alice was holding gradually turned into a pig before her eyes; finally, we may see an object moving and changing shape at the same time (as does a walking mammal). Using the experimental technique of "apparent movement" (which, interestingly, has been shown to occur in some animal species [16]), the conditions under which these perceptions are elicited can be empirically investigated.

Ullman's project is to discover the series of computations that the visual system performs on the input-pairs so as to arrive at an interpretation of the (2-D) array in terms of (3-D) replacement, motion, or change. In particular, he asks whether (and how) these distinct percepts can be differentially generated without assuming reliance on high-level concepts of specific 3-D objects (such as fish or sticklebacks), and even without assuming the prior recognition of a specific overall shape (such as a sort of narrow pointed ellipse with sharp projections on its upper surface). As the ethologist might put it, Ullman attempts to follow Lloyd Morgan's Canon, in asking what are the minimal computational processes that need to be posited to explain motion-perception. As we shall see, Ullman is misled by his concentration on mathematically minimal computations into assuming a specific hypothesis which is ethologically implausible--but this does not destroy the general interest of his approach.

As regards the visual interpretation of each array considered in isolation, Ullman relies on the work of David Marr, who has studied the information picked up from the ambient light by the retina, and the image-forming computations performed on it by peripheral levels of the visual system [17]. The first stage of visual computation, according to Marr, is the formation of a "Primal Sketch," an image consisting of descriptions of the scene in terms of features like shading-edge, extended-edge, line, and blob (which vary as to fuzziness, contrast, lightness, position, orientation, size, and termination points). These epistemological primitives are the putative result of preprocessing of the original intensity array at the retinal level—that is, they are not computations performed by the visual cortex (still less, the cerebral cortex). Marr defines further computations on these primitive descriptions, which group lines, points, and blobs together in various ways, resulting in the separatiopn of figure and ground. He stresses that these perceptual computations construct the image, which is a symbolic description (or articulated representation) of the scene based on the initial stimulus-array. The computations are thus interpretative processes, carried out by the visual system considered as a symbol-manipulating system rather than simply as a physical transducer (though Marr attempts to ground his computational hypotheses in specific facts of visual psychophysiology).

Starting with Marr's basic meaningful units, Ullman defines further visual computations which would enable the system, presented with two differing views, to make a perceptual decision between replacement, motion, or change. Ullman divides the computational problem faced by the visual system into two logically distinct parts, which he calls the correspondence and the interpretation problems. (The latter term unfortunately obscures the fact that all these computations, including Marr's, are interpretative processes, carried out by the visual system in its role as a symbol-manipulating device.)

The correspondence problem is to identify specific portions of the changing image as representing the same object at different times. This identity-computation must succeed if the final perception is to be that of a single object, whether in motion or in change. Conversely, the perception of replacement presupposes that no such identity could be established at the correspondence stage. The interpretation problem is to identify parts of the input arrays as representing objects, with certain 3-D shapes, and moving through 3-space (if they are moving) in a specific way. In principle, correspondence- and interpretation-computations together can distinguish between the three types of perception in question. And, if specific hypothetical examples of such computations are to be of any interest to students of biological organisms, they should be able to distinguish reliably (though not necessarily infallibly) between equivalent changes in the real-world environment.

This last point is relevant to the way in which Ullman defines specific correspondence- and interpretation-algorithms. In principle, any part of one 2-D view could correspond with (be an appearance of the same object as) many different parts of another; similarly, any 2-D view has indefinitely many possible 3-D interpretations. (Anyone who doubts this should recall the images facing them in distorting mirrors at funfairs.) Faced with this difficulty, Ullman makes specific assumptions about normal viewing conditions, and takes into account certain physical and geometrical properties of the real world, as well as (human) psychological evidence based on studies of apparent motion. Accordingly, he formulates a hypothetical set of computational constraints which he claims will both assess the degree of match between two views so as to choose the better one, and typically force a 3-D interpretation which is both unique and veridical. For instance, for the correspondence stage he defines "affinity functions" that compute the degree of match between two points or short line-segments, depending on their distance, brightness, retinal position, inter-stimulus (time) interval, length, and orientation. And for the interpretation stage, he defines a way of computing the shape and motion of a rigid object from three views of it, making his system assume that if such a computation succeeds then it is indeed faced with a rigid body in motion (as opposed to two different objects or one object changing its shape). He justifies this by proving mathematically that, except in highly abnormal viewing conditions, three views of a rigid object can uniquely determine its shape and motion.

Given that Ullman's computations can indeed interpret correspondence, shape, and motion in a wide range of paired 2-D views (which can be tested by running his system in its programmed form on a computer provided with the relevant input), what are we to say about the ethological importance of his work?

The first thing to notice is that Ullman embodies implicit assumptions about the physics and geometry of the real world, and about biologically normal viewing conditions, into the computations carried out by the visual system. It is plausible that many species may have evolved such implicit computational constraints. That is, the animal's mind may implicitly embody knowledge about its external environment, which knowledge is used by it in its perceptual interpretations. Something of the sort seems to be true for migratory birds, who have some practical grasp of the earth's magnetic field or of stellar constellations; and, as I shall suggest presently, the kingfisher may have some practical grasp of the refractive properties of water.

What is ethologically implausible about Ullman's hypotheses is not that they embody some knowledge about material objects and normal viewing conditions, but rather that they assume the perception of rigid objects to be basic, while perception of non-rigid movement is taken to be a more complex special case. Mathematically, of course, the perception of non-rigid motion is more complex; but this does not prove that it is biologically secondary to the perception of rigid objects. At least in the higher animals, it is more likely that the visual perception of shape and motion have evolved in response to such biologically significant environmental features as the gait or stance of hunter or prey, or the facial grimaces and tail-waving of conspecifics. The fact that human beings do not always perceive the correct (rigid) structure when presented with a mathematically adequate though impoverished stimulus, may be due not (as Ullman suggests) to their failing to pick up all of the mathematically necessary information in the stimulus, but rather to their using computational strategies evolved for the perception of non-rigid objects which— even when directed at rigid objects—need more information than is present in the experimental stimulus concerned C18D. Admittedly, a robot could be provided with an Ullmanesque capacity to perceive rigid objects in motion; but whether any creature on the phylogenetic scale employs such visual mechanisms is another question.

Our friend the kingfisher apparently possesses computational mechanisms which can discover the real position of a fish at varying depths in the water. Ullman's general approach suggests that these could well be relatively low-level processes, not requiring cerebral computations (as puzzling out Snell's law presumably does). For the visual computations algorithmically defined by Ullman do not depend on high-level processes capable of identifying (recognizing) objects as members of a specific class: the system does not need to know that an object is a fish, or even that it has the 3-D shape that it has, in order to know that it is an object. Nor does it need any familiarity with the object; that is, it does not need to have experienced those two views in association beforehand. Ullman therefore suggests (contra empiricists and Piaget) that a baby—or, one might add, a kingfisher—can see that two appearances are views of one and the same object even if it has never seen that sort of object before, and even if it has no tactile or manipulative evidence suggesting that they pertain to one and the same thing. These conclusions follow from the fact that all of the correspondence-computation, and much of the interpretation-computation, is via low-level, autonomous processes that do not depend on recognition of the input as a familiar 3-D object. The correspondence-computations match primitive elements (those defined by Marr) in successive views, and do not depend on computation of the overall shape as a whole.

It follows that creatures incapable of computing shape in any detail, or of recognizing different classes of physical object, may nonetheless be able to compute motion. As the example of von Uexkull's sea-urchin suggests, this is no news to ethologists, who often have behavioural evidence that an animal can perceive motion though they doubt its ability to be aware of detailed shapes. But Ullman's achievement is to have complemented this empirically-based intuition by a set of admirably clear hypotheses about precisely what visual computations may be involved, at least in the human case. That some of his hypotheses are biologically dubious does not destroy the ethological interest of his general approach.

Ullman's work also casts some light on our kingfisher-cartoon. For if the general shape, the location, and the motion of objects can be computed in a low-level, autonomous fashion, then it is not impossible that a kingfisher may possess comparable perceptual mechanisms capable of computing the depth of a fish in water. The refractive index of water would be implicitly embodied in these computational mechanisms, perhaps in an unalterable fashion. So a kingfisher experimentally required to dive into oil might starve to death, like newborn chicks provided with distorting goggles that shift the light five degrees to the right, who never learn to peck for grains of corn in the right place [19]. This assumes (what is the case for the chicks), that the kingfisher utilizes an inborn visuomotor coordination, linking the perceptual and active aspects of its Umwelt, a coordination that is not only innate but unalterable. Psychological experiments on human beings, and comparable studies of chimps, show that these species by contrast can learn to adjust to some systematic distortions of the physics of the visual field [20].

In their paper asking whether chimps are lay psychologists, Premack and Woodruff remark in an aside that one might also enquire whether they are lay physicists. Before being in a position to do this at any level of detail, one will need a clearer sense of what the content of a lay physics might possibly be. The foregoing discussion of Ullman's work suggests some part of the answer, in articulating assumptions about the physics and geometry of 3-D objects viewed in air (and, perhaps, in water) that may inform the *Umwelten* at least of some animals. But, presumably, human beings and many other species possess many more concepts and inferential structures that embody everyday knowledge of the material world, much of which knowledge may be acquired through learning. Some recent work in A.I., which admittedly is programmatic rather than programmed, is an interesting preliminary attack on this problem.

In his "Naive Physics Manifesto," P. J. Hayes asks how one might construct a formalization of our everyday knowledge of the physical world C211. Ethologists may be tempted to dismiss such an enquiry as irrelevant to their problems: human beings have Newton and Einstein, whereas animals do not, so human knowledge of physics cannot be relevant to enquiries about chimps, beavers, or bees. That this would be an inappropriate objection is evident from the fact that the *Punch* cartoon I mentioned earlier would have been almost as funny if it had figured a human fisherman rather than a kingfisher. Not only do we not usually think of Snell's law when we try to net a fish or tickle a trout, but we could not use it to help us do so even if we did. Similarly, we do not balance a bicycle by applying the formulae of mathematical dynamics. Our everyday intuitions of concepts such as *weight*, *support*, *velocity*, *height*, *inside/outside*, *next to*, *boundary*, *path*, *entrance*, *obstacle*, *fluid*, and *cause* (to name but a few) are pretheoretical. It is this pretheoretical knowledge which interests Hayes.

It is apparent that some animals share much of this pretheoretical knowledge with us—often, as in the case of the kingfisher, also knowing things which we do not. (Though in some cases "pretheoretical knowledge" may be grounded in a small number of independent condition-action rules, corresponding broadly to Gibson's notion of perceptual "affordances," rather than in prelinguistic conceptual networks of the sort posited by Hayes.) A cat or monkey leaping from wall to wall, or branch to branch, needs some representation of *support* and *stability*, and diving animals need some grasp of the difference between *solids* and *fluids*, as well as of *depth*, *movement*, and *distance*. Chimps clearly have some grasp of notions such as *inside*, *obstacle*, *place* . . . We will not be in a position to ascertain how much grasp, of which concepts, until we are clearer about the nature of these concepts in our own case. And this means knowing the perceptual evidence in which the concepts are anchored and the motor activities which test for them or which are carried out on the basis of conditional tests defined in terms of them. For example, newborn creatures who refuse to cross a "visual cliff" apparently have some innate procedure for recognizing the absence of *support*, where the object to be supported is their own body. It does not follow that they understand in any sense that the bottom bricks of a tower support the top ones—although this is something which a leaping animal living in a jungle or an untidy house may have to learn. To understand a concept involves having some representation of the inferences that can usefully be drawn to link it with other concepts in the same general domain. *Support*, for instance, has something to do with *above* for leaping creatures who can recognize the potential for action in a pile of bricks. Hayes outlines some ways in which the core concepts of naive physics, and groups of cognate concepts, may be organized, so that inferential paths can be traced between them. His work is an intriguing beginning of a very important enterprise, which should help us understand *how* perceptual experience functions in the control of motor action.

A word of warning is in order here, however. Hayes is primarily interested in the human *Umwelt*, which is informed through and through by natural language. It is true that our earliest knowledge of naive physics is prelinguistic: the baby's sensorimotor understanding is prior to her acquisition of English or French. But it follows from Hayes' account of meaning that, once such natural languages are acquired, the meaning of the more primitive core concepts is altered--not merely added to. In principle, then, even if we had a precise account of adult human knowledge of *inside*, *support*, and *behind*, we could not equate any part of this with the chimp's knowledge simply by jettisoning those parts of it influenced by our linguistic representations. Rather, we would need to be able to trace the development of our naive physical concepts, distinguishing their earlier, sensorimotor, forms from the later, linguistically-informed, semantic contents and inferential patterns. Hayes makes some relevant remarks, but even more apposite here is the computationally-informed work of the psycholinguists G. A. Miller and P. N. Johnson-Laird, who have studied the basic perceptual procedures in which our linguistic abilities are grounded [22].

Miller and Johnson-Laird define a number of perceptual discriminations in detailed procedural terms, utilizing what is known about our sensorimotor equipment and development. They then show how these discriminatory procedures could come to function as the semantic anchoring of our lexicon. For example, perceptual predicates that can be procedurally defined include the following spatial descriptions: *x* is higher than *y*; the distance from *x* to *y* is zero; *x* is in front of the moving object *y*; *y* is between *x* and *z*; *x* has boundary *y*; *x* is convex; *x* is changing shape; *x* has the exterior surface *y*; *x* is included spatially in *y*; *x*, *y*, and *z* lie in a straight line; *x* travels along the path *p*. They give both psychological and physiological evidence for the primacy of these notions, and they use them to define object-recognizing routines of increasing power. Their sensitivity to computational issues leads them to ask not only *which* predicates are involved in a certain judgment, but *when* each predicate is applied in the judgmental process. (For example, the logically equivalent "*y* over *x*" and "*x* under *y*" are not psychologically equivalent: the first term in the relation should designate the thing whose location is to be determined, while the second should represent the immobile landmark that can be used to determine it.) The perceptual routines they define as the meaning of words such as "in," "on," "outside," and "at" are surprisingly complex.

Were a chimp to grasp the meaning of "in" or "on" in Ameslan, therefore, this would presuppose extremely complex perceptual computations on the chimp's part. And animals which, unlike chimps, have no great manipulative ability, would not be able to compute those perceptual discriminations requiring motor activities such as putting bananas inside boxes, so that their understanding of naive physics would be correspondingly impoverished. Von Holst's studies of reafference [23], and Hein and Held's experiments on visual development in kittens [24], suggested that many perceptual discriminations require active bodily movement: insofar as this is so, the creature could not substitute an understanding of "putting in" derived merely from watching others. (It is perhaps worth remarking that limbless thalidomide babies apparently reach a normal understanding of physical concepts: whether their natural language plays an essential part in enabling them to do so is not known.) Irrespective of chimps' potential mastery of Ameslan, the implication common to the work of Ullman, Marr, Hayes, and Miller and Johnson-Laird is that the perceptual and motor abilities of animals far lower in the phylogenetic scale than chimps must be based on representational competences of a highly complex kind. So an increased sensitivity to computational issues might help ethologists to investigate the symbol-manipulations carried out by different species, and to compare Umwelten in a systematic fashion.

In addition to empirical observations (about which, more in the following section), it may be that general results in the abstract theory of computation might help in this systematic comparison. If it could be shown, for example, that a given type of representation in principle could not express a certain type of information, or that it would be computationally enormously less efficient than some other type of representation, such insights might help guide the ethologist in attributing specific representational capacities to different animals.

For instance, abstract considerations show that computational mechanisms of a certain type (namely, "perceptrons," of which an example would be a nervous net with no significant prior structure) simply cannot achieve specific kinds of learning or spatial pattern recognition [25]. Since it is abundantly clear that animal brains do have a significant prior structure, this result is somewhat academic from the point of view of the ethologist. But other results of this general type might be more relevant. For example, in discussing what F. Rosenblatt had termed "perceptrons," M. L. Minsky and S. Papert claimed to show that certain mechanisms capable of performing some nontrivial computations are incapable of performing others which at first sight might appear to be within their range [26]. Perceptrons are parallel-processing devices which make decisions on the basis of weighted evidence from many local operators, and various physiological examples have been suggested by cybernetically-inclined neurophysiologists interested in pattern-recognition and "self-organizing systems." Minsky and Papert sought to show, by way of abstract considerations alone, that no simple perceptron (without loops or feedback paths) could compute spatial connectedness, though it could compute convexity. Similarly, they claimed that no system without significant prior structure could in practice learn discriminations of high complexity, even given the existence of feedback paths.

Clearly, results such as these are relevant to the representational capacity of nervous systems of different kinds, whether in the form of more or less complex nervous nets or of highly structured cerebral systems. Whether these abstract considerations can soon be brought into articulation with specific neurophysiological data is another question, since in only very few cases can we realistically hope to have an adequate (still less, complete) understanding of the neural connections within an entire nervous system.

Another suggestive example of abstract work that might throw light on issues of interest to ethologists is provided by John McCarthy. McCarthy has long been interested in the representation of basic epistemological concepts (such as those discussed by his student, Hayes), and has recently embarked on what he terms "meta-epistemology," the attempt to define general representational or computational constraints on the sorts of mechanisms in principle capable of grasping particular notions [27]. (The account of perceptrons was in fact an early example of meta-epistemology, but was not conceived as part of an integrated research-programme directed to a wide range of representational systems.)

Again, A. Sloman has shown that "analogical" representations may be in various ways more computationally efficient than "Fregean" ones [28]. He defines an analogical representation as one in which there is some significant correspondence between the structure of the representation and the structure of the thing represented. By contrast, a Fregean representation need have no such correspondence, since the structure of the representation reflects not the structure of the thing itself, but the structure of the procedure (thought process) by which that thing is identified. To understand a Fregean representation is to know how to interpret it so as to establish what it is referring to, basically by the method described by the logician Frege as applying functions to arguments. Analogical representations, however, are understood or interpreted by matching the two structures concerned (that is, of the representation itself and of the domain represented), and their associated inference-procedures, in a systematic way. Applying this distinction to our kingfisher cartoon, for example, the formula expressing Snell's law is a Fregean representation, whereas the diagram itself (with the lines representing the paths of light and constructing the relevant angles) is an analogical representation.

An example of the use--and usefulness--of analogical representation has been provided by B.V.Funt, who has followed Sloman's suggestions by programming a system that can reason from visual diagrams [29]. Funt utlizes the 2D space inherent in the hardware of the machine as an analogue of 2D paper, so that a diagram is embodied in the machine as a certain state of a 2D visual array, or "retina." The system's task, given a diagram like that of Figure 2,

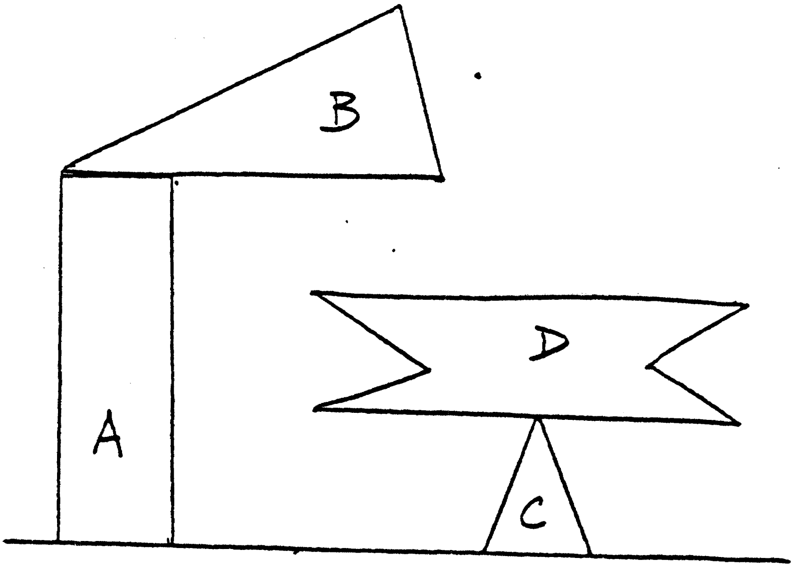

FIGURE 2.

is to discover whether the arrangement of blocks depicted is stable, and--if it is not--to predict the movements (falling, sliding, motion ended by contact with another block) and the final state of the various blocks. The answers to these questions are discovered from the diagram (given certain simple diagrammatic transformations carried out by the system, which are structurally analogous to changes that would happen in the real world), rather than being computed in terms of abstract mathematical equations and specific numerical values.

Previous programs that could recognize stability or instability of putative block-structures did so by computing sophisticated equations of physics and quantitative parameters, and one such program (one of the planning programs previously mentioned) needed over 80% of its computational resources for these calculations alone [30]. What is more, those programs were unable to predict the specific structural changes that would follow on an instability. But, much as it is "obvious" to us from the diagram (though not from a verbal or mathematical description of the same state-of-affairs) that B will hit D, that D will then tilt with its left half moving downwards, and that B will end up touching both A and D but <u>not</u> the ground, so it is easily discoverable by Funt's program that this is what will happen.

Briefly, the program "imagines" gradual changes in the position of the blocks by exploiting the 2D nature of the retina in which the diagram is embodied. So for instance it imagines gradually moving an unstable block (such as B) downwards, pivoting on the relevant point of support. It studies "snapshots" of the successive positions, and so discovers specific points of contact with coincidentally present blocks (such as D) which will interrupt the fall that would have been predicted by a theoretical physicist from equations and measurements describing A and B. As in this case, many detailed relations between blocks are implicit in the diagrammatic representation which could be explicitly stated only with the greatest difficulty. To take another example of this advantage of the diagrammatic representation, consider the recognition of empty space. What space is initially empty, and what would remain empty after stabilization of the blocks, can be directly discovered from the diagram and the imagined snapshots. But previous "blocks-world" programs have had to rely on highly counterintuitive assumptions about empty space, and/or have had to make complex mathematical or logical calculations to <u>deduce</u> the empty space in the scene.

Funt's work is relevant to the topic of naive physics discussed earlier. He points out that the physical knowledge exploited by his system is comparable to that of the lay person rather than the physicist. Thus the system has simple computational procedures, or "perceptual primitives," which address the visual array in parallel so as to identify area, centre, point of contact, symmetry, and so on. These spatial notions are likely to be useful in many different problem domains. Also, the program has knowledge of <u>qualitative</u> physical principles relevant to its actual tasks, such as that if an object sticks out too far it will fall, and that it will pivot around the support point nearest to the centre of gravity. Moreover, since it is able to discover the empty space, and also those spaces that would remain empty throughout stabilization changes, it possesses a type of knowledge that would be crucial to an animal looking for a pathway or for a safe space through which to move. Leaping animals, at least those whose weight might cause significant changes in the terrain leapt upon, presumably have some understanding of support and of potentially dangerous or unstable structures. For instance, chameleons clambering in trees seem to be capable of making a number of these judgments, preferring thick branches to thin ones and avoiding blind ends or gaps. Experimental study might help show what types of instability various animals are able to recognize, and perhaps whether they are able to distinguish any class of scenes as the likely outcome of a specific sort of instability. Are they able, for example, to distinguish between unstable structures differentially likely to collapse onto a baby animal underneath?

Funt has shown that Sloman's distinction between types of representation can be exemplified in computational terms, but much as Sloman's work is suggestive rather than definitive, so Funt's work is exploratory only. Among its specific limitations, Funt mentions its total ignorance of velocities, acceleration, and momentum. He remarks that were these matters to be included in a future version of the program (which of course would enable types of prediction currently impossible to it), they would have to be represented in terms of equations. But it is not obvious that some useful qualitative distinction between <u>fast</u> and <u>slow</u> might not be available to some creatures incapable of formulating equations. Ethological evidence is in principle relevant to this question, but so also would be an abstract understanding of the computational power of Fregean and analogical representations. If principled results were to be arrived at within computational logic, expressing the advantages and disadvantages of these representational modes, we might be better able to understand the cognitive potential available at different points in the phylogenetic scale.

<u>IV</u>: <u>Problems of Experimental Validation</u>

When contrasting planning programs and production systems in Section III, I pointed out that both these different approaches can model behaviour that is apparently controlled by some overall representation of the task. This is a special case of the general truth that there is always, in principle, more than one computational model capable of matching observed behaviour. So an ethologist who had produced a computational model of the diving kingfisher, for instance, could not thereby be certain of having captured the bird's psychology. Indeed, this was implicitly recognized by the <u>Punch</u> cartoonist: <u>if</u> the bird were consciously applying Snell's law, its dives would be (as they are) appropriately placed--but it does not follow that this is in fact the explanation of its diving ability. However, this caveat is itself a special case of the even more general truth that <u>any</u> scientific theory is necessarily underdetermined by the evidence. That this underdetermination causes methodological problems is well-known to every practising scientist. Were ethologists to produce computational theories, then, they woul;d be no worse off on this account than any other psychologist faced with the task of testing theory against data.

The special difficulty is not how to choose between several alternative computational accounts, once we have got them, but how to arrive even at <u>one</u> in the first place. People unfamiliar with A.I. typically underestimate the procedural-representational complexity underlying action and perception, and may even be unaware that there are unsolved computational problems related to everyday descriptions of behaviour. That is, behavioural descriptions are assumed to be unproblematic which in fact are deeply puzzling. The existence of various interpretative and representational capacities is taken for granted by most ethologists, who concentrate on asking which of these capacities are shared by which species. The A.I.-worker, by contrast, is primarily interested in how such capacities are computationally possible.

For instance, Premack and Rumbaugh ask whether chimps can do things which humans can do. Can a chimp interpret a movie as representing a second individual trying to solve a problem, like reaching bananas or switching-on a heater? Can a chimp plan ahead of time, either on its own behalf or on behalf of its fellow? Can two chimps cooperate in the solution of a task, perhaps using artificial symbols as publicly observable indicators of the tool that is required at a given stage of the problem? And so on .... But the computationally-inclined psychologist is more likely to ask, and to be primarily interested in, how these things can be done, irrespective of which species manage to do them. How is it possible for a creature to be a "lay psychologist," able to ascribe specific intentions, beliefs, and difficulties to another individual? How is it possible for a creature to form means-end plans for reaching a desired object, plans within which other objects are represented as instruments to the overall end? How is it possible for an external symbol, as well as one in the internal representational medium of the creature's mind, to be employed by one animal and recognized by another as a request for a specific tool? How is it possible for a creature to perceive apparent movement, or to distinguish visually between replacement, motion, and change? It is this difference in theoretical focus which has led one A.I.-worker to acknowledge the "fascination" of ethologists' studies of chimps, for example, and yet to complain that such studies are premature:

> In the long run we shall all learn more if we spend a little less time collecting new curiosities and a little more time pondering the deeper questions. The best method I know of is to explore attempts to design working systems [i.e., programs] that display the abilities we are trying to understand. Later, when we have a better idea of what the important theoretical problems are, we'll need to supplement this kind of research with more empirical studies [31].

For the ethologist who is interested in "comparative psychology," in the question of what achievements different animal species are capable of, this complaint will fall wide of the mark. One may legitimately be interested to discover the limits of what chimps, beavers, and bees can do. Many such questions have remained unasked by professional ethologists, because of the inhibitory influence of behaviourism--and even of the founding fathers of ethology, who were anxious to avoid sentimental anthropomorphism. But it remains true that a deep understanding of animals' abilities, which would carry us from "natural history" to "psychological theory," will require careful attention to the computational processes underlying these observed abilities. And computational ideas, in the meantime, may sometimes help to suggest specific empirical questions about the structure of behaviour which the etholgists might otherwise have left unasked--some examples were mentioned in Section III, with reference to planning programs and chimps' problem-solving. In addition, computational ideas will be helpful in framing questions about the evolution of representational and procedural powers. For instance, a deeper understanding of the computational processes required for vision in the higher mammals might throw light on the evolution of specifically linguistic competence [32].

V: A.I. and Animal Consciousness

Anyone who imagines (as many do) that because of its computerized, technological, base A.I. can have nothing useful to say about consciousness is mistaken. For instance, we have already seen that Ullman throws some light on the differences between three phenomenologically distinct types of experience--none of which, in cases of "apparent movement," constitute veridical perceptions of the real world. Whether his account of the generation of these experiences is correct is not the main issue here. The point is that he has offered a theoretical account of differential phenomenology that can be empirically investigated, and which if correct would explain how and why these distinct experiences arise when they do. Similarly Marr, on whose work Ullman draws, claims that the Primal Sketch--which is computationally defined, not introspectively posited--is very close to the image of which we are consciously aware.

To the extent that there are reasons for thinking animals to be capable of the same computations as ourselves, there is reason for positing analogous experiences in them. Some such evidence is behavioural/psychological: there is experimental evidence for example that apparent movement is experienced by some animal species--from the praying mantis, through guppies, to guinea pigs [33]. Or rather, perhaps one should say that all these species have been shown to possess visual mechanisms which cannot distinguish between smooth movement and abrupt changes of position, whereas only the higher animals (such as guinea pigs) can plausibly be assumed to have a phenomenal life in something like the sense in which we understand this term [34]. Some is physiological: if an animal has a peripheral visual physiology similar to ours, then its peripheral processing may be similar also. And some is computational: if programmed models of computational processes predict an animal's behaviour as well as they do that of human beings, then there is reason to suspect that it may enjoy comparable experiences.

Of course, we need to be very careful here, and to prove ourselves faithful disciples of Lloyd Morgan. Even in the human case, perceptual discriminations may be carried out without resulting in any conscious phenomenology, as the example of "blind sight" shows [35]. We have seen that human sensorimotor concepts of <u>support</u> and <u>inside</u> may be radically altered (not merely added to) by linguistically mediated understanding. And Marr admits that the visual image of which we are aware is commonly somewhat different from the Primal Sketch, which is normally enriched and further articulated by higher-level computational processes working "top-down" on the array. For example, to see a tree as an <u>oak</u> (rather than an <u>ash</u>, or a <u>chestnut</u>....) requires that the high-level concept of an oak-tree be applied to the Primal Sketch, which itself will feature a rounded-triangular green blob, rather than an oak-tree. In general, any phenomenological distinction requiring linguistic representations in its generation cannot arise in animals lacking natural language. For example, the human *Umwelt* includes many emotional dimensions which only a Proustian mastery of language can clearly express, and which cannot be attributed to animals. The subtle differences between reproach, resentment, and censure, for instance, rest on cognitive distinctions that can be attributed only in embryonic form (if at all) to non-language-using species.

Often, it is not easy to distinguish (in the human case) those experiences which are conscious merely in the sense of being open to self-report (which pain is) from those which arise only as a result of self-conscious examination, or deliberate reflexive computation, by the person concerned (which pain presumably does not C36D). Many psychologists have remarked that consciousness arises at points of difficulty in the execution of motor tasks, and that it presumably helps focus attention on the detailed adjustments required. Whether it follows that a chimp or a dog, apparently carefully carrying out some task or executing some plan, is similarly conscious at points of difficulty is unclear.

In the absence of introspective verbal report, we are on shaky ground in attributing particular conscious experiences to animals—or to human babies. This drawback of animal psychology is not removed by a computational approach. But the example of apparent movement shows that consciousness, or phenomenal experience, is not in principle outside the theoretical scope of this approach. Were we to understand human consciousness better, we would be better placed to compare it to animal phenomenology. "Understanding consciousness better" would involve a grasp of how different experiences are generated, what computational roles they play in the scheduling of attention in the mind as a whole, and how they influence motor activity. These are all notoriously difficult questions, but they fall within the computational (symbol-manipulating) domain rather than outside it.

There is another notorious difficulty which is more intractable. This is the basic philosophical conundrum of how a phenomenal experience can possibly arise from an assemblage of material stuff (whether the stuff be protoplasm or anything else). I think it is not impossible that computational insights may eventually help us to a satisfactory answer to this question, though no such answer is yet available. But in this, the computational approach is no worse off than other forms of theoretical psycholgoy. Psychologists in general, including ethologists, commonly leave this obscure philosophical problem on one side, assuming—at least for the human case—that consciousness *does* exist, and that there are interesting theoretical questions to be asked about its structure, generation, and psychological conditions. The example of Ullman's work should suffice to show that these questions can usefully be cast in a computational mould.

VI: Postscript

The moral of this paper is not so much that A.I.-workers have achieved solutions which should be adopted by cognitive ethologists, as that A.I.-workers are asking questions whose answers--when they are achieved--can hardly fail to be of interest to ethologists. Drawing this moral reminds me that an author who recently urged the use of computational concepts in neurolinguistics was chided by a critic for "the crackling of promissory notes" that pervaded his paper [37]. I am only too aware that promissory notes have been liberally issued in the preceding pages, so much so as to threaten to deafen us with their crackling. But new ideas in psychology can, after all, be no more than promising.

NOTES

1. T. S. Collett & M. F. Land, "How Hoverflies Compute Interception Courses," J. Comp. Physiol., 125 (1978), 191-204.

2. Aaron Sloman, "The Primacy of Non-Communicative Language," in M. McCafferty & K. Gray, eds., The Analysis of Meaning (ASLIB and Brit. Comp. Soc., 1979).

3. Jacob von Uexkull, "A Stroll Through the World of Animals and Men," in C. H. Schiller, ed., Instinctive Behavior: The Development of a Modern Concept (New York: International Univ. Press, 1957), pp. 5-82. (Orig., 1934.)

4. I have discussed purposive action and relevant theories at length in M. A. Boden, Purposive Explanation in Psychology (Cambridge, Mass.: Harvard Univ. Press, 1972; Brighton, Sussex: Harvester Press, 1978).

5. Konrad Lorenz, "The Nature of Instinct: The Conception of Instinctive Behaviour," in Schiller, ed., Instinctive Behavior, pp. 129-175. Quote from p. 157. (Orig., 1937.)

6. D. S. Lehrman, "The Physiological Basis of Parental Feeding Behavior in the Ring Dove," Behaviour, 7 (1955), 241-286; "Effect of Female Sex Hormones on Incubation Behavior in the Ring Dove," and "Induction of Broodiness by Participation in Courtship and Nest-Building in the Ring Dove," J. Comp. Physiol. Psychol., 51 (1958), 142-145, 32-36.

7. David Premack & Guy Woodruff, "Does the Chimpanzee Have a Theory of Mind?,"; D. R. Griffin, "Prospects for a Cognitive Ethology,"; and E. S. Savage-Rumbaugh et al., "Linguistically Mediated Tool Use and Exchange by Chimpanzee,"--all in Behavioral and Brain Sciences, 1 (1978), No.4.

8. See Chapter 12 of M. A. Boden, Artificial Intelligence and Natural

Man (New York: Basic Books; Hassocks, Sussex: Harvester Press, 1977), for a description of many such programs.

9. Konrad Lorenz, *Behind the Mirror: A Search for a Natural History of Human Knowledge* (London: Methuen, 1977).

10. See note 8, above, for a general account. Some primary sources include: R. E. Fikes, P. E. Hart, & N. J. Nilsson, "Learning and Executing Generalized Robot Plans," *Artificial Intelligence*, 3 (1972), 251-288; E. D. Sacerdoti, "Planning in a Hierarchy of Abstraction Spaces," *Artificial Intelligence*, 5 (1974), 115-136; S. E. Fahlmann, "A Planning System for Robot Construction Tasks," *Artificial Intelligence*, 5 (1974), 1-50; G. J. Sussman, *A Computer Model of Skill Acquisition* (New York: Elsevier, 1975); R. Power, *A Model of Conversation*, Working Paper, Experimental Psychology Lab., University of Sussex, 1976.

11. R. Davis & J. King, "An Overview of Production Systems," in E. W. Elcock & D. M. Michie, eds., *Machine Intelligence 8* (Mew York: Wiley, 1976).

12. R. M. Young, *Seriation by Children: An Artificial Intelligence Analysis of a Piagetian Task* (Basel: Birkhauser, 1976).

13. R. M. Young, *Mixtures of Strategies in Structurally Adaptive Production Systems: Examples from Seriation and Subtraction*, (D.A.I. Research Report No.33, Dept. A.I., Univ. Edinburgh, 1977).

14. Shimon Ullman, *The Interpretation of Visual Motion* (Cambridge, Mass.: M.I.T. Press, 1979); Shimon Ullman, "The Interpretation of Structure from Motion," *Proc. Royal Soc. London*, 203 (1979), 405-426.

15. J. J. Gibson, *The Perception of the Visual World* (Boston: Houghton-Mifflin, 1950); J. J. Gibson, *The Senses Considered as Perceptual Systems* (Ithaca, N.Y.: Cornell Univ. Press, 1966); J. J. Gibson, *The Ecological Approach to Visual Perception* (N.Y.: Houghton-Mifflin, 1979).

16. I. Rock, E. S. Tauber, & D. P. Heller, "Perception of Stroboscopic Movement: Evidence for its Innate Basis," *Science*, 147 (1964), 1050-1052; and *Science*, 153 (1966), 382. Also K. U. Smith, "The Effect of Partial and Complete Decortication upon the Extinction of Optical Nystagmus," *J. Gen. Psychol.*, *25* (*1941*), *3*-*18*.

17. David Marr, "Early Processing of Visual Information," *Philos. Trans. Royal Soc.*, 275 (1976), 483-524; David Marr, "Representing Visual Information," *Lectures on Mathematics in the Life Sciences, Vol, 10: Some Mathematical Questions in Biology* (1978), pp. 101-180; David Marr, Visual Information Processing: The Structure and Creation of Visual Representations," *Sixth Int. Joint Conf. Artif. Intell.* (1979), pp. 1108-1126.

18. This point is made by Aaron Sloman, "What Kind of Indirect Process is Visual Perception?," in *Behavioral and Brain Sciences* (in press: reply to S. Ullman's "Against Direct Perception," *ibid*.)

19. E. H. Hess, "Space Perception in the Chick," *Scientific American*,

195 (1956), 71-80.

20. G. M. Stratton, "Some Preliminary Experiments on Vision," *Psychol. Rev.*, 3 (1896), 611; "Vision Without Inversion of the Retinal Image," *Psychol. Rev.*, 4 (1897), 341; I. Kohler, "Experiments with Goggles," *Scientific American*, 206 (1962), 62.

21. P. J. Hayes, "The Naive Physics Manifesto," in Donald Michie, ed., *Expert Systems in the Micro-Electronic Age* (Edinburgh: Edinburgh Univ. Press, 1979), pp. 242-270.

22. G. A. Miller & P. N. Johnson-Laird, *Language and Perception* (Cambridge, Mass.: Belknap Press, 1976).

23. E. von Holst, "Relations Between the Central Neervous System and Peripheral Organs," *Brit. J. Animal Behaviour*, 2 (1954), 89-94.

24. R. Held & A. Hein, "Movement-Produced Stimulation in the Development of Visually Guided Behaviour," *J. Comp. Physiol. Psychol.*, 56 (1963), 872.

25. M. L. Minsky & Seymour Papert, *Perceptrons: An Introduction to Computational Geometry* (Cambridge, Mass.: M.I.T. Press, 1969).

26. Frank Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psych. Rev.*, 65 (1958), 386-408.

27. John McCarthy, personal communication. See also his "First-Order Theories of Individual Concepts and Propositions, in Michie, ed., *Expert Systems*, and John McCarthy & P. J. Hayes, "Some Philosophical Problems frm the Standpoint of Artificial Intelligence," in Bernard Meltzer & Donald Michie, eds., *Machine Intelligence 4* (Edinburgh: Edinburgh Univ. Press, 1969), pp. 463-502.

28. Aaron Sloman, "Intuition and Analogical Reasoning," in his *The Computer Revolution in Philosophy: Philosophy, Science, and Models of Mind* (Hassocks, Sussex: Harvester Press, 1978), pp. 144-176.

29. B. V. Funt, "Problem-Solving with Diagrammatic Representations," *Artificial Intelligence*, 13 (1980), 201-230.

30. Fahlmann's "*BUILD*," cited in note 10, above.

31. Aaron Sloman, "What About Their Internal Languages?," *Behavioral and Brain Sciences*, 1 (1978), 602-603.

32. R. L. Gregory, "The Grammar of Vision," in R. L. Gregory, *Concepts and Mechanisms of Perception* (London: Duckworth, 1974), pp. 622-629. *Cf.* also E. H. Lenneberg, *Biological Foundations of Language* (New York: Wiley, 1967).

33. See references cited in note 16, above.

34. For a provocative account of the philosophical difficulties involved in attributing phenomenal experiences to other species, see T. Nagel, "What is it Like to Be a Bat?," *Philosophical Review*,

(1974), 435-451.

35. Larry Weiskrantz, "Trying to Bridge Some Neurological Gaps Between Monkey and Man," *Brit. J. Psychol.*, 68 (1977), 431-435.

36. D. C. Dennett, "Why You Can't Make a Computer That Feels Pain," in *^s *Brainstorms; Philosophical Essays on Mind and Psychology* (Brighton, Sussex: Harvester Press, 1979), pp. 190-232.

37. Howard Gardner, "Commentary on Arbib & Caplan, 'Neurolinguistics Must be Computational'," *Behavioral and Brain Sciences,* 2 (1979), p. 465.

oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo