The Role and Nature of Learning for an Artifact

Joanna Bryson

April 11, 1994

The purpose of this paper is to discuss the nature of behavior acquisition for a robot, specifically for Cog, the humanoid robot currently under development by Brooks and Stein et al. here at MIT. Using the term “learning” to describe this acquisition is dangerous, since many people have come to associate the word with specific kinds of processes. Within the context of this paper, I use *learning* to mean any process which enables an agent to acquire a new behavior which it can perform autonomously. This definition is deliberately broad enough to include such things as hardware design. If this seems strange, think about how it relates to expressions like “Evolution has learned to give birds of prey hooked beaks and talons.” For Cog, the challenge is to begin with an agent with no behavior (and no body) and end with something that behaves like a human, or at least a child. Cog has a lot to learn.

With Cog, the decision has already been made to make the agent a robot, and to make it humanoid. Why? One reason is that we *know*, by biological example, that human behavior can be achieved using such an architecture. We *suspect* that many capabilities that aren't typically thought of as physically based, such as language comprehension, might be better achieved by a being that shares our perception of the physical world in as many ways as possible(Brooks & Stein 1993). Recent speculation suggests that these ways need to include not only perception gained through passive observation, but through shared experiences in coping with and manipulating the environment. For example, Mark Johnson suggests that a concept like “path”, which we learn initially by physically moving to goal locations, becomes a fundamental metaphor from which we derive meaning from all sorts of situations(Johnson 1987). For example, we come to understand the events of maturation as a journey on a path, then we understand careers in terms of maturation.

From the beginning then, Cog will have certain fundamental behaviors "built in". For example, it will receive optic information from a 180° cone of space in front of it's head. It will fall over. Can these acts be considered components of intelligent behavior? My answer is "yes". Consider that neither rocks nor plants can *do* anything to cause themselves to fall — in the natural world, the ability to make yourself fall over would be considered to be an act of intelligence, though not much. By inclination, humans find even plants that can move (such as the venus fly-trap) to be intelligent.

I bring these behaviors up not just for further establishing my semantics, but as examples of the kinds of components of which intelligence consists. Artificial Intelligence, and before it Philosophy of Mind, have historically considered intelligence to be the ability to logically manipulate symbols that represent things that have meaning within our environment(Newell & Simon 1963). Hand in hand with this notion has been the idea that learning consists of general mechanisms for deriving knowledge from experience in a similar, logical way. The falling behavior is modular and specific to the nature of the robot's joints and the action of gravity, and it is not deliberate. Therefore it does not conform to the traditional *formal* definition of intelligence.

The model of intelligence as general reasoning has persisted despite the fact that nature provided clear examples of intelligence that could not have been learned by our conventional understanding of the term. Animals are born knowing how to swim, walk, find the ocean. They can develop complex skills without apparent training, such as animals that learn to fly or hunt when raised in isolation from their species by humans. Without an alternative theory of intelligence, the "knowledge sets" considered necessary to perform these activities were classified as "instincts", something qualitatively different from human thought.

These distinctions have begun to break down, for two reasons. One is that as evolution has replaced religious explanations of the origin of intelligence, it has seemed less likely that there would be such distinct differences between animals and humans, though the fact that only humans seem capable of complex language is strong evidence for at least some intellectual specialization. The other reason is the increasing evidence that *people* are born with many specialized capabilities and biases. For example, there are certain "universal" rules of grammar that any natural language, that is a language spoken by people from birth, obey. When adults invent a new language, for example a pidgin created by speakers of different languages needing to communicate, they may disobey these rules. But their children, who learn the language from them, will reintroduce the universal grammar laws to the new language(Bickerton 1987). Compare this to the animals that learn their natural behavior with only the ignorant encouragement of human foster parents.

If human and animal intelligence are related, then understanding animal behavior acquisition may help us develop strategies for learning in Cog. Ethology tries to understand animal behavior as an evolutionary derivative, just like animal morphology. Traditionally, it was believed that animals learn only through a general process of being able to create associations.

> The [general process assumption] position is that all learning is based on the capacity to form associations; there are general laws of learning that apply equally to all domains of stimuli, responses, and reinforcers; the more frequent the pairings between the elements to be associated, the stronger the associative strength; the more proximate the members of an association pair, the more likely the learning(Gallistel, Brown, Carey, Gelman & Keil 1991).

This learning by association is also called "conditioning", and it does appear to be a general learning mechanism that follows various parameter-based rules across species. However, animals *cannot* learn to associate just any stimulus with any response. For example pigeons can learn to peck for food, but cannot learn to peck to avoid a shock. They can, however, learn to flap their wings to avoid a shock.(Hineline & Rachlin 1969). Originally, these results were thought to be the result of some kind of constraint placed on general learning. But recently, some researchers have come to believe the opposite — that the *capacity* to be conditioned is only present for a privileged set of stimuli and responses. For example, rats are presented with "bad" water with two separate cues for its badness: a funny taste, and flashing lights or buzzing. The water is made bad in one of two ways — either it is poisoned and makes them sick later, or they are shocked as a consequence of drinking it. Rats who are shocked learn to avoid the noise and lights, but not the funny taste. Rats that are poisoned avoid the flavor and not the noise and lights(Garcia & Koelling 1966). This indicates rats are using survival-oriented learning mechanisms — poison is often indicated by smell or taste, while acute pain is often the consequence of something that can be seen or heard.

Galliste et. al use these experiments as evidence for their hypotheses that learning by an individual organism serves as a last resort for evolution. It is only relied on when there is no other way to determine the behavior, because the competence involved requires flexibility on a less than evolutionary time scale. They give other examples of learning they do not consider associative, such as the precise acoustic location of prey by barn owls (necessary because it is dependent on the shape of the individual bird's head) and bees learning the ephemeris of the sun for navigation (dependent on season and latitude). These are not cases of classical conditioning. The animals seem to be born with a limited number of variables which are instantiated at a certain stage of development by simple observation of the world.

An example of this that comes very close to humans is in the vervet monkeys.(Seyfarth, Cheney & Marler 1980) These monkeys have 3 distinctive cries for predators which warn their troop to take appropriate defensive action. These cries are dedicated to pythons, martial eagles, and leopards. Baby vervets make cries from a very early age, but across overly general objects. For example, they may give the "eagle" cry for anything in the sky, the "leopard" cry for any animal, the "python" cry for a stick on the ground. Thus they are born attending to the sorts of stimuli they need to be aware of, but they learn fine discrimination experientially.

Gallistel et al. do not claim that there are no general learning mechanisms; in fact, they describe two kinds of generalization. First, there are some cases when a single form of learning can be applied to multiple tasks. They suggest that the same mechanism used for localizing oneself within one's environment can be used for modeling the foraging availability within one's range (foraging animals spend time foraging in different locations proportionally to the relative amount of fodder). Second, they admit the existence of general learning laws, such as some of the equations describing classical conditioning. They suggest these laws indicate underlying evolved mechanisms for learning.

Of course, from a biological point of view, there can be no doubt there are "shared mechanisms". To pull back again briefly from the problems of learning on the individual organismic level, it is clear that many species of animals share common morphologies. To the extent that behavior can be defined as the interaction between an agent and its environment, and that the physical build of creature will have a large impact on that interaction, clearly these species share learned behaviors. Roboticists have learned that building a robot with these physical-interaction aspects of behavior in mind can greatly reduce the amount of computational complexity needed to get a robot to perform a given task.(Hallam 1991) As an extremely simple example, to avoid having a robot get stuck in holes in its environment make it (or it's avoidance sensory perimeter) too large to fall into them. Many species of animals have used such size adaptation to take advantage of specific niches.

On the individual level again, one obvious common biological mechanism for learning is the neural system. An animal's neural system controls and communicates its behaviors and perceptions. Neural systems are composed of neurons. Neurons change in two ways over the lifetime of an animal. First, they grow, and sometimes die. Growth serves to connect the neuron to possible stimulators or receivers. Death can eliminate connections that prove not to be useful. Much neural growth is probably roughly predetermined, like other physical systems. Some complex mappings between sensory regions seem to be learned without external "training" or reenforcement, like the barn owls or the bees learned above(Kohonen 1982).

The second way a neuron can change is by strengthening or weakening the amount of impact a stimulus it receives from one specific input location (dendrite) has on the output of the node. These inputs are received from adjoining neurons, and the place where they communicate is called a synapse. When "synaptic learning" was discovered there was a great deal of excitement, because it became a strong candidate for explaining how general learning, or learning by association, could take place. For a simple example, if a particular input is frequently associated with a neuron firing, then that synapse connection will be strengthened. Synapses may either serve to excite or inhibit a particular neuron.

In the 1950's computer scientists were intrigued by the possibility of doing computation on massive parallel arrays of extremely simple processors modeled on the neuron. This research started with Rosenbloom's perceptron work and was met with great success as it seemed finally possible to allow a computer to train itself (by doing essentially "synaptic learning" — affecting the impact of inputs on outputs for various nodes). But research fizzled in the late sixties when it was realized that some fairly basic classes of inputs (such as xor) could not be sorted using known training methods. New strategies were developed in the early eighties, and many classification applications were found for neural nets, such as hand writing recognition. However, as a model of general cognition, they are again beginning to fade from popularity as complex problems (like language) remain recalcitrant.

This failure is hardly surprising, since neural nets are really simply tools for discriminating inputs — that is, association machines. To assume a neural net can display full intelligent behavior would be to assume that *all* learning is based entirely on a simple mapping of a set of inputs to the correct outputs. This paper has already taken a far different stance. However, *some* learning is of this type. Also, it seems likely not only from biological architecture, but from performance considerations that neural nets are a better model of natural intelligence than the old serial-processing logic theories. For example, neural nets are good at the kinds of tasks we are good at, like recognizing things very quickly (in very few sequential steps); bad at the things we are bad at, like math and remembering sequential events (both of these require us to use props like paper and pencil if they are at all complex), and show similar tolerance for damage (if any part of a logic program is lost the entire result tends to be invalid — people and neural nets can lose many processors and show just a degradation of performance).

Precisely at this point, damage and failures, the *discrepancy* between neural nets and human brains becomes apparent. Human brains are not a single, homogeneous network. Not only are there biological differences between different sections of the brain, but damage to different areas results in different kinds of behavioral deficits. (McCarthy & Warrington 1988)

Many people are now introducing more complex models for the task of an individual neuron. These theories often suggest that a single neuron might represent a feature detector (Poggio 1990) or a coincidence detector(MacKay 1985). To MacKay, a single neuron might represent a hypothesis about the world — the hypothesis that the effect of the output of that neuron is the correct reaction to the world state the sensory organs it connects to are reflecting. When significant coincidences are detected, that is, the hypothesis is proved true, the output effect of this neuron is strengthened and it becomes part of the organism's "knowledge" about the world. Poggio, on the other hand, describes cells that are explicitly dedicated to recognition of particular features, by virtue of evolutionary processes including their location within the hierarchy of brain cells. These cells train to recognize the feature they are needed for fairly quickly, and because the value they return is not discrete but rather responds as a Gaussian function to related input, relatively few dedicated feature detectors are needed for any particular object or concept. Poggio is motivated by the computational need for "lookup table" efficiency in the brains actions. Because neural acts are chemically based, relatively few computational steps can possibly happen in the amount of time we use to select our behaviors. Both Poggio and MacKay's notions are supported by research evidence that certain individual cells do seem to respond to specific sensory input, for example an individual face(Perrett, Mistlin & Chitty 1987). Whether we believe that either of these explanations could explain all of cognition, both mechanisms sound useful for individual adaptations to a flexible environment.

Recently within robotics there has been a movement towards developing intelligence in a way something like the end product of MacKay's proposal. Reactive, behavior-based models of control create robot behavior by factoring the desired behavior into modular parts. Each module reacts directly as an augmented finite state machine to its inputs, which may be either sensory input or the monitored outputs of other, more low-level modules(Brooks 1991). To date, these methods have mostly been applied to small mobile robots. On the level of individual learning, these robots create their own complex path behaviors in finding their way around in a dynamic environment. Some can even find and retain systems of landmarks for navigation using neural networks. But most of the behaviors are strictly emergent from the interaction of the simple reactive modules that were engineered into them, with each other and with the robot's environment.

Ethologists and philosophers of mind have also become more interested in a modular concept of intelligence. One issue that particularly interests them is the "compilation" of a sequence of actions into a reflex-like *skill*(Karmiloff-Smith 1992). For example, when you are learning tennis, you learn a sequence of behaviors including where to place your feet, how high to toss the ball, when to hit it. When you become expert at this task, you stop thinking about each individual component of it and think instead of serving as a single act. The same applies to driving a car, typing, or grasping a cup. Although we are not born knowing how to do these tasks, once we become expert at them we can best model them as single reflexive acts, rather than a long sequence of actions.

The forms of learning I have been describing so far can be broken into three classes:

1. provided learning,

2. required learning, and
3. open learning.

For example, in the neurology discussion, the motor neurons (with predetermined growth) are provided learning, the forming of topological sensor maps in early development is required learning, and the changes that occur at the synapses as a consequence of experience are open learning. Before returning to discussing learning within the specific context of Cog, the humanoid robot, I will give a more formal description of these classifications and their possible mechanisms.

Provided learning is the behaviors determined without any action by the intelligent agent itself. For animals, this includes the bodies they are born with, their reflexes, and some extent of what might be called "instinctive behavior". For robots, this is their physical hardware, and any software behaviors that have no variable storage content, that is, that are not influenced by the agent's own previous experience but only by its direct current environment.

Required learning is domain-specific, preordained learning performed by the individual agent that is required for its normal "adult" behavior. Examples from biology include things like the barn owl acoustic localization explained above, bird song learning, and bees learning the ephemeris of the sun for navigation. In robotics, examples include self-calibration of sensors or preliminary training of recognition networks. This kind of learning can often be thought of in terms of parameter adjustment. Often it occurs in animals during specific periods of their early development, while in robots it might occur before the the robot is considered fully functional. Some "recalibration" sorts of required learning persist through the whole lifetime of the agent.

Open learning is the closest equivalent to the traditional definition of learning — it is freer, more general learning and is likely to occur over the entire lifetime of the agent. Examples of this are the associative learning and the skill compilation strategies mentioned above. Notice that this is still not necessarily the kind of complete, general, logical learning long proposed as the complete story. Not only does it coexist with, and in most cases play a subordinate role to, the afore mentioned forms of learning, but it may well be constrained to specific domains, as in the examples of the rats and pigeons from Gallistel above.

Notice that each of these classifications could be described in terms of its predecessor. The mechanisms for required learning are necessarily part of the provided learning array. And open learning could be seen as a specific example of required learning — one required by some animals in order to live within their complex evolutionary niches.

Particularly in biology, the border lines between classifications may be unclear in some cases, like the physical growth of the organism. In a robot, however, these distinctions are clear because of we can easily distinguish the mechanisms that underlie them. Provided learning will never change over the lifetime of the robot. Required learning adjusts specific parameters in narrow, well defined ways, using special purpose mechanisms. Open learning uses general techniques to add new skills, new reactions. Where the organic ambiguity is reintroduced is in the engineers' decision of which behaviors should fall into which class of learning.

Now is a particularly hard time for engineers to use biology as a guidline for making decisions about mind architecture. The most recent research into human cognition has done more to prove that the brain is modularized in ways we don't expect than to present a coherent explanation of what those ways are(McCarthy & Warrington 1988). Even the fundamental model of modular intelligence is non-intuitive and controversial(Dennett & Kinsbourne 1992). Nevertheless, I will use the reactive, behavior-based model proposed by Brooks as the basic paradigm for discussing Cog's intelligence. Presumably, it will be under this methodology that the provided learning will be accomplished for Cog.

One of the basic tenants of behavior-based cognition is that complex behavior can arise from the interaction of a small number of simple actions. This has been demonstrated true in the activities of simple robots and programs that simulate natural behavior using this model(Brooks 1991). It can even be seen in the everyday world, for example in the complex harmonic and rhythmic consequences of children singing "rounds". The question is whether the scalability of emergence is adequate to allow us to achieve human-like behavior on Cog. We are depending not only on our ability to select the right base behaviors, but that evolution has only needed or been able to develop and maintain a sufficiently small number of behaviors that we will be capable of engineering and understanding something on a similar scale.

One of the freedoms allowed us by the new learnings in neuropsychology is that we can cluster and reuse our behaviors in ways that have nothing to do with our traditional classifications and intuitions about how we see and react to the world. For example, coordinating activities between the different modules within the agent can be viewed as a multi-agent or cooperative problem, not only by the engineers but by the agent itself. Cog could use the same modules to learn about controlling and coordinating its own actions as its physical and social interactions with the external world.

In the case of human infants, one of the early tasks facing them is distinguishing aspects of themselves from their environment. We know infants initially show ignorance of their own manipulations despite somatic feedback — they upset themselves by getting their hands caught on each other or their own hair. How do they begin this classification? One possibility is by detecting the coincidence between some act of activation on their own part, and the achievement of some goal. But in the early months these goals are most reliably met indirectly — it is easier to signal a desire to a care giver than to fulfill a desire oneself. Psychologists know that an individual's competence at social interaction is highly involved with that person's self image, and also that one's self image and one's image of one's parents are intimately linked. This could also explain why somatic input (other than the reward system) is given such low priority in solving the inappropriate grasping quandaries above.

Cog will have two arms to learn to control. It will need to learn basic behaviors for each — for example, grasping an object. It might seem that the same behaviors could be used for both hands, but if the object is only sized for a one hand grasp, and Cog wanted it, what would prevent both hands from colliding in their attempt to grasp it? A very simple form of cooperation would simply be hand dominance. If the object triggers a single hand grasp, the right hand's grasp might be supressed by the left hand's. But if some other module has already supressed grasping in the left hand, the right hand would make the gesture.

But what if a person is working with Cog? What if the person moves to grasp the object first? A traditional approach would be to treat this as a seperate case, as being the same as the object being occluded in some other way. But from a reward-oriented point of view these situations could quite possibly be identicle, in fact it could well be that the human will have an easier time reaching the object and putting it in Cog's hand than Cog would. So perhaps solving both these problems with visual feedback would be more behaviorally efficient. It would also be more humanoid, since in children handedness takes many months to become apparent, and may be partly a factor of preferentially dedicating training to the hand that emerges as slightly more coordinated.

What other specific recommendations can we make about learning in Cog? For one, following as a rule of thumb Gallistel's hypothesis about individual learning being a methodology of last resort would probably be a good idea. It would not be easy to give Cog the two years a baby has before it begins combining words to see how Cog's capacities develop, because the cycle time on failure would be too long. Our learning mechanisms are likely to be both more primitive (in terms of specificity and robustness to noise) and more complex (in terms of possible failure points) than a child's. Additionally, there will be the problem of dealing with less consistent hardware and sensory input.

At the same time, we need to be careful about making oversimplifying assumptions. For example, it might seem that one kind of learning Cog would not need to deal with that children do is physical growth. Consequently, it might be tempting to hardcode behaviors like two handed manipulation in terms of Cog's known dimensions. This overlooks two problems. One is that while Cog's dimensions *in space* may not change, other dimensions might. For example, strength. In the case of humans, infants are physically incapable of lifting anything with which they can injure themselves by dropping the object on theirself. Similarly, infants cannot swing anything heavy enough hard enough to injure themselves (either by bludgeoning or by joint strain) or their caretakers. In the early stages of Cog we will probably want to build similar constraints into the system for the safety of both the robot and its operators. However, eventually people will probably want humanoid robots with strength comparable to a human's. Suddenly giving this increased strength to a new revision of Cog would be little safer than giving it to the first one. Instead, we would probably want to find a way to allow the manipulation to adapt while the strength is scaled up. In addition, incidental fluctuations in strength are quite likely, due to hardware considerations.

The other problem with programming too many "hard-wired" manipulations into Cog's repertoire is that the mechanisms we use to learn object manipulation might be fundamental to other desirable human capabilities, such as language. Although never experiencing the physical scaling humans do already creates a difference — perhaps metaphors concerning the equivalence of power and size will be harder to grasp — this is not a reason to compound the problem[1].

On the other hand, there are probably other fundamental stages we can bypass. For example, one thing we will need to have early, as either required or provided learning, are reward systems. These are used in most forms of open learning. Even if we do not use rewards directly, it will be important for Cog to be *perceived* as reacting positively to a reward in order to encourage human interaction with it. Humans have rewards associated with fulfilling biological needs: they are happy when they are being fed while hungry, they are unhappy if they are too cold. But we also have other rewards that are less physical: a smiling or frowning face, a kind or a mean voice. Some of these rewards may well have components that are learned, possibly by association with need fulfillment. For example, it is important that an infant learn to recognize and prefer the comfort of its mother's voice over that of a stranger's, though it is also useful to be able to tell when it is pleasing strangers. However, the universality of some basic expressions makes it quite likely these are examples of required learning, with some amount of template built in.

For Cog, we will probably be safe building reward functions in based immediately on these verbal cues as the fundamentals, rather than trying to learn them by association. In fact, to the extent possible we should probably build rewards in as provided learning, or where needed in required learning performed outside normal operation using supervised training. Examples of reward systems Cog could use are the afore mentioned social feedback (happy and unhappy faces and voices), negative reaction to pressure on touch sensors over a certain weight, and positive reaction to motion or bright colors in its field of view. The social goals will allow people to train Cog, both intentionally and unintentionally. The touch anti-goal would serve as a form of pain, to prevent Cog from damaging itself or its environment. The visual

[1] Besides, my intuition as a developer is that manipulating will be more crucial to human-like behavior than physical growth is, though I could be very wrong. One aspect of different scale is that it does clearly differentiate people more or less into age groups, so it could be helpful to a sense of identity and the importance of peers. However, people who actually retain inappropriate scale (such as dwarves and midgets) often report if they grow up in isolation from others with the same situation that they are shocked when they first see others with their syndrome because they don't envision themselves as being that different.

goals will encourage Cog to observe potentially educational situations in the world.

Free learning will be necessary for Cog to be convincingly humanoid. If Cog manages to demonstrate a behavior or sequence of behaviors that please itself either directly or through feedback from a human observer, we would like to think some mechanism made it more likely these behaviors would be repeated. This would quite likely be similar to the "skill compiler" mentioned in Karmiloff-Smith[2]. Quite likely a mechanism for conditioning will also prove necessary. For example, it would be nice if Cog came to associate behaviors that were consistently rewarded as rewards in themselves. It could then "entertain itself" and develop skills with, say, blocks, without requiring the presence or at least the forced adulation of its supervisors.

The purpose of this paper has been to discuss a model for behavior acquisition. The hope is that it will be useful in helping us think about engineering the behavior of Cog. By exploring relevant research in human, animal, and robot intelligence I have tried to both clarify and justify my position, and finally I have made some specific recommendations, though necessarily general and high level, for the design of Cog. These suggestions are really more examples of the processes of learning than any kind of blueprint — I do not think they necessarily address the most crucial behaviors of Cog.

The main point of this paper is that learning can take place in three ways: without any action on the part of the individual (provided learning), as a simple and logically necessary consequence of the provided behaviors (required learning), or by active general learning processes (open learning). The challenge for the engineers is to decompose the problem of human-like intelligence into behaviors introduced by one of these mechanisms, and to ensure these behaviors can interact in such a way that the desired overall behaviors can emerge.

[2]Notice that acquiring expertise is often considered the transformation of a skill from from *declarative* knowledge to *implicit* knowledge. But I am purposely avoiding any discussion of the utility of "declarative" knowledge in Cog, partly because I don't think we will need it for these infancy stages, and partly because it is too closely linked with the notion of consciousness in many people's minds. I will simply point out that many non-verbal animals are capable of learning complex sequences of behavior. I would speculate that there is probably a mechanism close to conditioning for learning short sequences, and that these sequences might then be recursively chained into something that looks much more complex.

References

Bickerton, D. (1987), *Language & Species*, The University of Chicago Press, Chicago, Illinois.

Brooks, R. A. (1991), Intelligence Without Reason, *in* 'Proceedings of the 1991 International Joint Conference on Artificial Intelligence', pp. 569–595.

Brooks, R. A. & Stein, L. A. (1993), Building Brains for Bodies, Memo 1439, Massachusetts Institute of TechnologyArtificial Intelligence Lab, Cambridge, Massachusetts.

Dennett, D. C. & Kinsbourne, M. (1992), 'Time and the Observer: The Where and When of Consciousness in the Brain', *Brain and Behavioral Sciences* **15**, 183–247.

Gallistel, C., Brown, A. L., Carey, S., Gelman, R. & Keil, F. C. (1991), Lessons From Animal Learning for the Study of Cognitive Development, *in* S. Carey & R. Gelman, eds, 'The Epigenesis of Mind', Lawrence Erlbaum, Hillsdale, New Jersey.

Garcia, J. & Koelling, R. A. (1966), 'The relation of cue to consequence in avoidance learning', *Psychonomic Science* **4**, 123–124.

Hallam, J. (1991), Autonomous Robots : From Dream to Reality, DAI Research Paper 526, University of Edinburgh, Cambridge, Massachusetts.

Hineline, P. & Rachlin, H. (1969), 'Escape and avoidance of shock by pigeons pecking a key', *Journal of Experimental Analysis of Behavior* **12**, 533–538.

Johnson, M. (1987), *The Body In The Mind*, University of Chicago Press, Chicago, Illinois.

Karmiloff-Smith, A. (1992), *Beyond Modularity*, MIT Press, Cambridge, Massachusetts.

Kohonen, T. (1982), 'Self-organized formation of topologically correct feature maps', *Biological Cybernetics* **43**, 56–59.

MacKay, D. M. (1985), The significance of 'feature sensitivity', *in* D. Rose & V. G. Dobson, eds, 'Models of the Visual Cortex', John Wiley & Sons Ltd.

McCarthy, R. A. & Warrington, E. K. (1988), 'Evidence for Modality-Specific Systems in the Brain', *Nature* **334**, 428–430.

Newell, A. & Simon, H. (1963), GPS: A Program that Simulates Human Thought, *in* E. Feigenbaum & J. Feldman, eds, 'Computers and Thought', McGraw-Hill, New York.

Perrett, D. I., Mistlin, A. J. & Chitty, A. J. (1987), 'Visual neurones responsive to faces.', *Trends in Neuroscience* **9**, 358–364.

Poggio, T. (1990), A Theory of How the Brain Might Work, Memo 1253, Massachusetts Institute of TechnologyArtificial Intelligence Lab, Cambridge, Massachusetts.

Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980), 'Monkey REsponses to three different alarm calls: Evidence of predator classification and semantic communication', *Science* **14**, 801–803.