MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A. I. Memo 899 **May, 1986**

ACHIEVING ARTIFICIAL INTELLIGENCE THROUGH BUILDING ROBOTS

Rodney A. Brooks

Abstract. We argue that generally accepted methodologies of Artifical Intelligence research are limited in the proportion of human level intelligence they can be expected to emulate. We argue that the currently accepted decompositions and static representations used in such research are wrong. We argue for a shift to a process based model, with a decomposition based on task achieving behaviors as the organizational principle. In particular we advocate building robotic insects.

Acknowledgments. This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the research is provided in part by an IBM Faculty Development Award, in part by a grant from the Systems Development Foundation, and in part by the Advanced Research Projects Agency under Office of Naval Research contract N00014 85 K 0124.

(c) Massachusetts Institute of Technology 1986

1. Introduction

There has been considerable philosophical debate on the possibility of "human level" artificial intelligence, centered around the notion that it requires as background the totality of practices which make up the human way of being in the world [Dreyfus 72, 86].

In this note we use a technical rather than philosophical argument that machines must indeed have a rich background of experience of being if they are to achieve human level intelligence. Unlike Dreyfus however, we conclude that artificially intelligent behavior is achievable with computers without the aid of holograms, resonance, or other holistic techniques. Rather, by adopting an incremental construction approach, progress towards this goal can be expected soon. (Naturally the author and his students are currently following this enlightened path.)

2. Three observations about the failure of AI.

In this section we give three reasons why AI has failed to live up to its early promises, and moreover why it will not live up to its current promises.

A. A lesson from evolution

We already have an existence proof of the possibility of intelligent entities–human beings. Additionally many animals are intelligent to some degree. (This is a subject of intense debate much of which really centers around a definition of intelligence.) They have evolved over the 4.6 billion year history of the earth.

It is instructive to reflect on the way in which earth–based biological evolution spent its time. Single cell entities arose out of the primordial soup roughly 3.5 billion years ago. A billion years passed before photosynthetic plants appeared. After almost another billion and a half years, around 550 million years ago, the first fish and vertebrates arrived, and then insects 450 million years ago. Then things started moving fast. Reptiles arrived 370 million years ago, followed by Dinosaurs at 330 and Mammals at 250 million years ago. The first primates appeared 120 million years ago and the immediate predecessors to the great apes a mere 18 million years ago. Man arrived in roughly his present form 2.5 million years ago. He invented agriculture a mere 19000 years ago, writing less than 5000 years ago and "expert" knowledge only over the last few hundred years.

This suggests that problem solving behavior, language, expert knowledge and application, reason, etc., are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. This part of intelligence is where evolution has concentrated its time—it is much harder.

[Lovejoy 81] has argued that human level intelligence was an extremely unlikely accident. The basis of his argument is that evolutionarily there is "no *a priori* advantage to intelligence, although it is a clear and unmistakable reproductive hazard" (increased encephalization does not in itself lead to dominance, e.g. only two species of *Proboscidea* remain and are survived by less-encephalized but equally large occupants of similar ecological zones—the hazard is the length of time necessarily devoted to parenting of offspring during the full development of the vastly more complex central nervous system). He argues

that evolution did not select for the majority of characteristics of the human brain in the context of intelligence. By and large the structures were already there in more primitive hominids. The evolution of habitual bipedal locomotion induced changes in the oral diaphram and in the nasopharyngeal passage. This, combined with changes induced in the oral cavity by greater seasonality of food sources, gave man the capacity to speak. He argues that this capacity reprogrammed the existing form of the brain into a symbol processing system at the higher levels and gave rise to cognition.

This supports my contention that mobility, acute vision and the ability to carry out survival related tasks in a dynamic environment provide a necessary basis for the development of true intelligence. [Moravec 84] argues this same case rather eloquently.

B. Abstraction as a dangerous weapon

Artificial Intelligence researchers are fond of pointing out that AI is often denied its rightful successes. The popular story goes that when nobody has any good idea of how to solve a particular sort of problem (e.g. playing chess) it is known as an AI problem. When an algorithm developed by AI researchers successfully tackles such a problem, however, AI detractors claim that since the problem was solvable by an algorithm, it wasn't really an AI problem after all.

Thus AI never has any successes.

But have you ever heard of an AI failure?

I claim that AI researchers are guilty of the same (self) deception. They partition the problems they work on into two components. The AI component, which they solve, and the non-AI component which they don't solve. Typically AI "succeeds" by defining the parts of the problem that are unsolved as not AI. The principal mechanism for this partitioning is abstraction. Its application is usually considered part of good science, not, as it is in fact used in AI, as a mechanism for self delusion. In AI, abstraction is usually used to factor out all aspects of perception and motor skills. I argue below that these are the hard problems solved by intelligent systems, and further that the shape of solutions to these problems constrains greatly the correct solutions of the small pieces of intelligence which remain.

Early work in AI concentrated in games, geometrical problems, symbolic algebra, theorem proving, and other formal systems (e.g. [Minsky 68], [Feigenbaum and Feldman 63]). In each case the semantics of the domains were fairly simple.

In the late sixties and early seventies the blocks world became a popular domain for AI research. It had a uniform and simple semantics. The key to success was to represent the state of the world completely and explicitly. Search techniques could then be used for planning within this well-understood world. Learning could also be done within the blocks world; there were only a few simple concepts worth learning and they could be captured by enumerating the set of subexpressions which must be contained in any formal description of a world including an instance of the concept. The blocks world was even used for vision research and mobile robotics, as it provided strong constraints on the perceptual processing necessary [Nilsson 84].

Eventually criticism surfaced that the blocks world was a "toy world" and that within it there were simple special purpose solutions to what should be considered more general problems. At the same time there was a funding crisis within AI (both in the US and

the UK, the two most active places for AI research at the time). AI researchers found themselves forced to become relevant. They moved into more complex domains, such as trip planning, going to a restaurant, medical diagnosis, etc.

Soon there was a new slogan: **"Good representation is the key to AI"** (e.g. *conceptually efficient programs* in [Bobrow and Brown 75]). The idea was that by representing only the pertinent facts explicitly, the semantics of a world (which on the surface was quite complex) were reduced to a simple closed system once again. Abstraction to only the relevant details thus simplified the problems.

Consider a chair for example. While the following two characterizations are true:

(CAN (SIT-ON PERSON CHAIR))

(CAN (STAND-ON PERSON CHAIR))

there is much more to the concept of a chair. Chairs have some flat (maybe) sitting place, with perhaps a back support. They have a range of possible sizes, requirements on strength, and a range of possibilities in shape. They often have some sort of covering material, unless they are made of wood, metal or plastic. They sometimes are soft in particular places. They can come from a range of possible styles. In particular the concept of what is a chair is hard to characterize simply. There is certainly no AI vision program which can find arbitrary chairs in arbitrary images; they can at best find one particular type of chair in carefully selected images.

This characterization however is perhaps the correct AI representation of solving certain problems; e.g., a person sitting on a chair in a room is hungry and can see a banana hanging from the ceiling just out of reach. Such problems are never posed to AI systems by showing them a photo of the scene. A person (even a young child) can make the right interpretation of the photo and suggest a plan of action. For AI planning systems however, the experimenter is required to abstract away most of the details to form a simple description in terms of atomic concepts such as PERSON, CHAIR and BANANAS.

But this abstraction is the essence of intelligence and the hard part of the problems being solved. Under the current scheme the abstraction is done by the researchers leaving little for the AI programs to do but search. A truly intelligent program would study the photograph, perform the abstraction and solve the problem.

The only input to most AI programs is a restricted set of simple assertions deduced from the real data by humans. The problems of recognition, spatial understanding, dealing with sensor noise, partial models, etc. are all ignored. These problems are relegated to the realm of input black boxes. Psychophysical evidence suggests they are all intimately tied up with the representation of the world used by an intelligent system.

There is no clean division between perception (abstraction) and reasoning in the real world. The brittleness of current AI systems attests to this fact. For example, MYCIN [Shortliffe 76] is an expert at diagnosing human bacterial infections, but it really has no model of what a human (or any living creature) is or how they work, or what are plausible things to happen to a human. If told that the aorta is ruptured and the patient is losing blood at the rate of a pint every minute, MYCIN will try to find a bacterial cause of the problem.

Thus, because we still perform all the abstractions for our programs, most AI work is still done in the blocks world. Now the blocks are slightly different shapes and colors, but their underlying semantics have not changed greatly.

It could be argued that performing this abstraction (perception) for AI programs is merely the normal reductionist use of abstraction common in all good science. The abstraction reduces the input data so that the program experiences the same perceptual world (*Merkwelt* [Uexküll 21]) as humans. Other (vision) researchers will independently fill in the details at some other time and place. I object to this on two grounds. First, as Uexküll and others have pointed out, each animal species, and clearly each robot species with their own distinctly non-human sensor suites, will have their own different *Merkwelt*. Second, the *Merkwelt* we humans provide our programs is based on our own introspection. It is by no means clear that such a *Merkwelt* is anything like what we actually use internally–it could just as easily be an output coding for communication purposes (e.g., most humans go through life never realizing they have a large blind spot almost in the center of their visual fields).

The first objection warns of the danger that reasoning strategies developed for the human-assumed *Merkwelt* may not be valid when real sensors and perception processing is used. The second objection says that even with human sensors and perception the *Merkwelt* may not be anything like that used by humans. In fact, it may be the case that our introspective descriptions of our internal representations are completely misleading and quite different from what we really use.

C. Studying epicycles

For most of the history of astronomy, an earth centered model was used. As observations became more accurate it was realized that the sun and planets did not have circular orbits around the earth. The solution was simple: Elaborate the assumed models. This led to the introduction of epicycles, smaller circular orbits whose centers rode on larger earth centered circles called deferents to explain perturbations to circular planetary orbits. This was good enough to explain the path in the sky traced out by the planets, but did not explain their variable speeds. Ptolemy introduced two further variants: he moved the centers of the deferents off the earth (i.e. made them eccentric) and then made the angular velocity of the centers of the epicycles constant relative to a point opposite the deferent center from the earth, known as the equant.

This scheme did not lend itself to a mechanically obvious construction of the universe and was therefore philosophically objectionable to some. Eventually Al-Tusi [Gingerich 86] produced a less objectionable model, without equants, with two more layers of epicyles on top of each of the old epicylces.

When the underlying model changed to a sun centered system all the centuries of scholarship and computation of epicycles went out the window. Gone. Useless.

I suspect much of current AI work will in hindsight appear to have been chasing details of epicycles. Current models for AI are based on concepts of knowledge and belief and symbol processing systems. I claim that all such things are concepts invented by observers of intelligent systems to explain them. The intelligent systems themselves do not work like that. They do not have atomic tokens as symbols. As observers we are ascribing tokens to complex classes of states of complex processes. I believe AI will progress significantly only when the fundamental underlying model being used moves towards process and away from state.

3. Intelligence from the bottom up.

The above considerations lead me to two major conclusions:

- *No toys.* There is great methodological danger in tackling AI problems in toy worlds. The definition of "toy" here includes all worlds where it is not up to the AI system itself to do all the understanding of the world itself without relying on a human interpreter. Likewise a world is a "toy" world if the AI system is not responsible for carrying out its actions in the world without a human agent to interpret its responses. Such requirements on an AI system thus force it to be part of a robot system acting in a *real world* for some definition of real. This is a much stronger definition of real world than is normally used.
- *Different modularity.* True intelligence requires a vast repertoire of background capabilities, experience and knowledge (however these terms may be defined). Such a system can not be designed and built as a single amorphous lump. It must have components. The same is true of existing AI systems. But true intelligence is such a complex thing that one can not expect the parts to be built separately, put together and have the whole thing work. We are in such a state of ignorance that it is unlikely we could make the right functional decomposition now. Instead we must develop a way of incrementally building intelligence. Simple systems need to be debugged in real worlds along the road to complex intelligent systems. A key problem then is how to perform a decomposition which leads to an incremental path to intelligence.

Worlds: toy and real

If the ultimate goal in building a human level intelligent system is to have it interact in some way with people in the world, then it should have as the basis of its experience and being, down to the lowest levels, that same world.

There has been a long tradition of building artificial worlds for mobile robots (e.g. [Nilsson 84], [Crowley 85], [Giralt et al 84]). Our earlier arguments were directed against such efforts which typically have constructed special rooms with walls and doorways having regular properties. Usually in such cases even the artifacts within the world were also specially constructed.

There is a more subtle mistake in building special environments however–often not realized by researchers who believe they are using a "natural" or "real" world (e.g. [Moravec 84]). All these past efforts in mobile robots have assumed a static environment. All methods of perception and planning are predicated in this fact. But any world involving people or even more than one robot is dynamic. Even an insect can survive (for a while at least) in a dynamic world populated by hostile humans. Any "real" world therefore should include dynamic aspects. Even better, the robot should be expected to operate in the presence of humans. It should operate in this manner even at its most primitive levels. Adding this capability later will turn out to be much more difficult; as a data point consider that no existing mobile robot project has been able to make this advance.

The best domain for trying to build a true artificially intelligent system is a mobile robot wandering around an environment which has not been specially structured for it.

Decomposition

There are many possible approaches to building an autonomous intelligent system. As with most engineering problems they all start by decomposing the problem into pieces, solving the subproblems for each piece, and then composing the solutions.

Typically in AI, the chosen decomposition is based on information processing functions. Thus, for example, there may be a data-base and an inference engine. Or, in more complex systems, there might be a stereo vision system producing a depth map, a model builder which turns that into three dimensional primitives and their relations, a map builder which integrates such observations into a global map, a task planner which uses the map information to plan a sequence of actions for a robot, and a task execution system which tries to carry out the plans.

Systems built under such a decomposition have a number of drawbacks. In particular the system is usually only as strong as its weakest link. In addition such systems usually have some central locus of control; perhaps procedural as in a central task planner, or perhaps declarative as in a blackboard. I think of the first as a central bottleneck through which even the most inconseqential actions must be planned. [Huttenlocher 84] has called the latter an error propagation mechanism. I propose that rather than slicing the problem on the basis of internal workings of the solution we should slice the problem on the basis of desired external manifestations of the intelligent system.

The idea is to decompose the desired intelligent behavior of a system into a collection of simpler behaviors. We build computational systems to achieve each of the more primitive behaviors and then compose them into a more complex system. Ideally each individual computational system should be independent of the others. In practice there will be some overlap (just as happens in a more usual decomposition where for instance it is hard to tell where the stereo depth system ends and the three dimensional model builder begins).

An additional question is how this decomposition into behaviors should be organized. In our work with mobile robots [Brooks 86] we have used two principles.

- Each behavior should achieve some task. I.e., there should be some observable phenomenon in the total behavior of the system which can be used by an outside observer to say whether the particular sub-behavior is operating successfully. As an example we have a *don't hit things* behavior for our robot. We refer to each sub-behavior as a task achieving behavior.
- A set of task achieving behaviors together provide the robot with some level of competence. They should be designed so that as new task achieving behaviors are added to the system, the level of competence increases. A level of competence is a somewhat informal specification of overall system performance. In [Brooks 86] we organized our task achieving behaviors in such a way as to generate strict linear ordering of levels of competence (see figure 1). We are currently working on a generalization of this structure.

Decomposing an intelligent system into independent task achieving behaviors has a number of positive benefits.

- There are many parallel paths of control through the system. Thus the performance of the system in a given situation is not dependent on the performace of the weakest link in that situation. Rather it is dependent on the strongest relevant behavior for

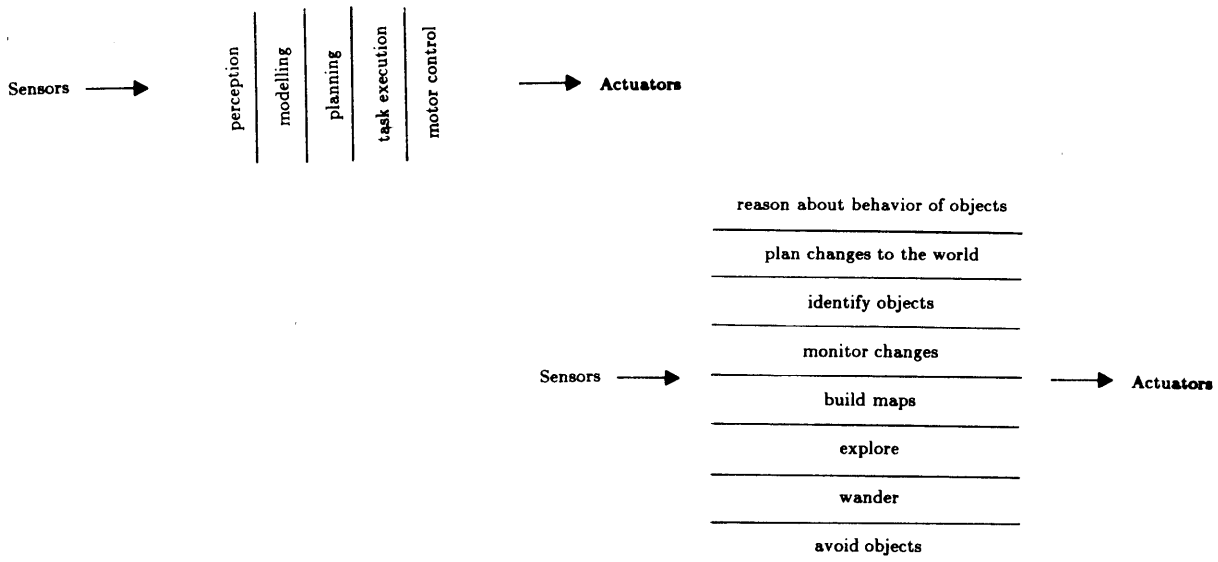

Figure 1. a. Functional decomposition, b. Behavior-based decomposition

the situation. That in turn is dependent on its own weakest link, but we are now minimizing over a set preselected for its strength!

- Often more than one behavior is appropriate in a given situation. The fact that the behaviors are generated by parallel systems provides redundancy and robustness to the overall system. There is no central bottleneck; procedural or declarative.
- With some discipline in structuring the decomposition, the individual task achieving behaviors can run on separate pieces of hardware. Thus it leads to a natural mapping of the complete intelligent system onto a parallel machine. The benefits are threefold; (1) redundancy again, (2) speedup, and (3) a naturally extensible system.

4. Artificial Insects

Insects are not usually thought of as intelligent. However they are very robust devices. They operate in a dynamic world, carrying out a number of complex tasks including hunting, eating, mating, nest building, and rearing of young. There may be rain, strong winds, predators, and variable food supplies all of which impair the insects' abilities to achieve its goals. Statistically however insects succeed. No human-built systems are remotely as reliable. Reliability concerns both mechanical systems and information processing systems. Thus I see insect level behavior as a noble goal for artificial intelligence practitioners. I believe it is closer to the ultimate right track than are the higher level goals now being pursued.

Progress

We have tested these ideas on how to decompose an intelligent system with a mobile robot operating in laboratory and machine room environments. It operates at insect level intel-

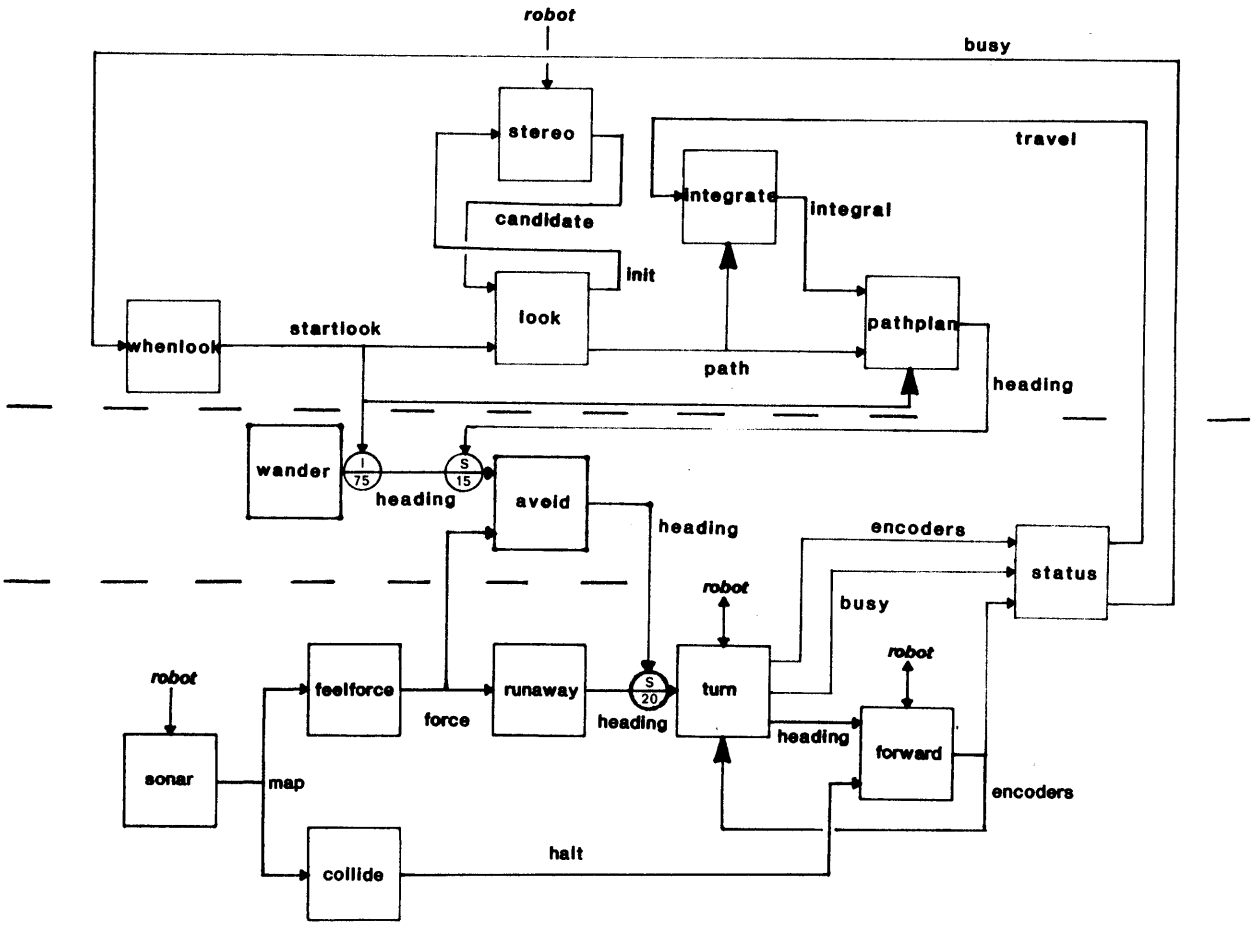

Figure 1. Three layered control system for an exploring robot

ligence. We have implemented the first three levels of task achieving behaviors shown in figure 1. Figure 2 shows the implementation of those three layers in terms of simpler computational modules. They operate in a dataflow fashion, sending messages asynchronously without guarantee of delivery or even acknowledgement of receipt. (We place such loose restrictions on the underlying computational engine in the belief that it will make it easy to construct a large extensible piece of hardware meeting these modest goals.)

With just the lowest level of control the robot uses sonar to find a large empty space and then sits there contented until a moving obstacle approaches. Two people together can successfully herd the robot just about anywhere–through doors or between rows of disk drives, for instance.

When the next level of control is added the robot is no longer content to sit in an open space. After a few seconds it heads off in a random direction. Our uncalibrated sonars and obstacle repulsion functions make it overshoot a little to locations where the **runaway** module reacts.

With the third level a sonar-based corridor finder (substituting for "stereo" in figure

2) usually finds the most distant point in the room. The robot heads off in that direction. People walking in front of the robot cause it to detour, but it still gets to the initially desired goal even when it involves squeezing between closely spaced obstacles. If the sonars are in error and a goal is selected beyond a wall, say, the robot usually ends up in a position where the attractive force of the goal equals the repulsive forces of the wall. At this point **avoid** does not issue any heading, as it would be for some trivial motion of the robot. The robot sits still, defeated by the obstacle. The **whenlook** module, however, notices that the robot is idle and initiates a new scan for another corridor of free space to follow.

Pragmatics

But wait! Pragmatism rears it ugly head. While artificial insects are fine and dandy from a purely scientific research point of view, is it plausible that funders will believe in their long range promise?

I claim that robot insects will gain more mass market acceptance than any expert system built with current technology. They have the potential to change our daily lives in much the way microprocessors have. Not in a single obvious way, but in many small incremental ways. For example microprocessors abound in the household environment and are sometimes the enabling technology of certain appliances. Microprocessors are under the hoods of our automobiles, in microwave ovens, on our wrists, in our compact disk players, in our television sets and VCRs and in our video games. There is a potential for even simple insect like artificially intelligent robots to play similar roles.

For example, suppose we want to build a vacuum cleaning robot. It will not be a robot that looks like a person and that carries the vacuum cleaner around; instead it will look like a vacuum cleaner, which has been modified so that under its own propulsion it wanders around the house and cleans the floor. Such a robot might well be insect–like in its intelligence level. But unfortunately it will be quite large and may not be able to get into the corners. So, there will be a natural niche in the home habitat for another little robot which will be truly insect like. I like to think of it as about two inches in diameter and about an inch thick. It will have four, perhaps six, legs sticking out of it. They will be piezo-electrically operated. It will not matter which way up this robot is—it can be thrown on the floor or fall down the stairs. It will sit on its belly and reach out its legs, digging them into the carpet to drag itself across the floor. It will be controlled by an architecture as we have described above and it will have a number of simultaneous goals. One thing it will like is corners. It will drag itself on its belly across the floor. (Incidentally it could be solar powered or powered from the ambient light, because it does not have to move fast. In fact there is an advantage because the cat will not be interested in it if it is slow moving.) It will use its legs as force sensors and thus detect when it has reached a wall. Once there, one of the other level goals takes over and it will start looking around for dirt. It will pick it up electrostatically (or with some as yet undetermined technology) and shove it in its belly somewhere. At the same time it will also be dragging itself along, but now in a different mode: it will be near a wall so it drags itself along the edge. When it gets to a tri-hedral vertex at the corner it will switch to another mode and sit there for a long time, picking up dirt. Then after a while it will get bored and shoot off in some random direction. In this way complete coverage of the house is achieved.

Now the only problem with this is, what should it do with the dirt? One approach is to have it go to some central location and get rid of the dirt. But that implies extremely good navigation abilities and a detailed model of the house. This is beyond a simple insect. Instead, there will be another layer of control utilizing sound sensors. It will listen for the big vacuum cleaner. When it hears it, it will run to the middle of the room and dump its guts all over the floor.

Such is the future.

Acknowledgements

Jon Connell, Peter Cudhea, Eric Grimson, Thomas Marrill, James Little and David Chapman made many helpful comments on earlier drafts of this paper.

References

[**Bobrow and Brown 75**] "Systematic Understanding: Synthesis, Analysis, and Contingent Knowledge in Specialized Understanding Systems", Robert J. Bobrow and John Seely Brown, in "Representation and Understanding", Bobrow and Collins eds, *Academic Press, New York*, 1975, 103–129.

[**Brooks 86**] "A Robust Layered Control System for a Mobile Robot", Rodney A. Brooks, *IEEE Journal of Robotics and Automation*, RA-2, April 1986.

[**Crowley 85**] "Navigation for an Intelligent Mobile Robot", James L. Crowley, *IEEE Journal of Robotics and Automation*, RA-1, March 1985, 31–41.

[**Dreyfus 72**] "What Computers Can't Do: A critique of Artificial Reason", Hubert Dreyfus, *Harper Row*, 1972.

[**Dreyfus 86**] "Why Computers May Never Think Like People", Hubert and Stuart Dreyfus, *Technology Review*, January 1986.

[**Feigenbaum and Feldman 63**] "Computers and Thought", E. A. Feigenbaum and J. Feldman eds, *McGraw-Hill, San Francisco*, 1963.

[**Gingerich 86**] "Islamic Astronomy", Owen Gingerich, *Scientific American*, April 1986, 74–83.

[**Giralt et al 84**] "An Integrated Navigation and Motion Control System for Autonomous Multisensory Mobile Robots", Georges Giralt, Raja Chatila, and Marc Vaisset, *Robotics Research 1, Brady and Paul eds, MIT Press*, 1984, 191–214.

[**Huttenlocher 84**] "Acoustic-Phonetic and Lexical Constraints in Word Recognition: Lexical Access Using Partial Information", *S.M. Thesis MIT Dept. EECS*, 1984.

[**Lovejoy 81**] "Evolution of Man and Its Implications for General Principles of the Evolution of Intelligent Life", C. Owen Lovejoy, in "Life in the Universe", edited by John Billingham, *MIT Press, Cambridge*, 1981, 317–329.

[**Moravec 83**] "The Stanford Cart and the CMU Rover", Hans P. Moravec, *Proceedings of the IEEE*, 71, July 1983, 872–884.

[**Moravec 84**] "Locomotion, Vision and Intelligence", Hans P. Moravec, *Robotics Research 1, Brady and Paul eds, MIT Press*, 1984, 215–224.

[**Minsky 68**] "Semantic Information Processing", Marvin L. Minsky ed, *MIT Press, Cambridge*, 1968.

[**Nilsson 84**] "Shakey the Robot", Nils J. Nilsson, *SRI AI Center Technical Note 323*, April 1984.

[**Shortliffe 76**] "MYCIN: Computer-based Medical Consultations", Edward H. Shortliffe, *Elsevier, New York*, 1976.

[**Uexküll 21**] "Ummwelt und Innenwelt der Tiere", J. Von Uexküll, *Berlin*, 1921.