MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1293 April, 1991

Intelligence Without Reason

Rodney A. Brooks

Prepared for *Computers and Thought*, IJCAI-91

Abstract

Computers and *Thought* are the two categories that together define Artificial Intelligence as a discipline. It is generally accepted that work in Artificial Intelligence over the last thirty years has had a strong influence on aspects of computer architectures. In this paper we also make the converse claim; that the state of computer architecture has been a strong influence on our models of thought. The Von Neumann model of computation has lead Artificial Intelligence in particular directions. Intelligence in biological systems is completely different. Recent work in behavior-based Artificial Intelligence has produced new models of intelligence that are much closer in spirit to biological systems. The non-Von Neumann computational models they use share many characteristics with biological computation.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for this research was provided in part by the University Research Initiative under Office of Naval Research contract N00014–86–K–0685, in part by the Advanced Research Projects Agency under Office of Naval Research contract N00014–85–K–0124, in part by the Hughes Artificial Intelligence Center, in part by Siemens Corporation, and in part by Mazda Corporation.

1 Introduction

Artificial Intelligence as a formal discipline has been around for a little over thirty years. The goals of individual practitioners vary and change over time. A reasonable characterization of the general field is that it is intended to make computers do things, that when done by people, are described as having indicated intelligence. Winston [**Winston 84**] characterizes the goals of Artificial Intelligence as both the construction of useful intelligent systems and the understanding of human intelligence.

There is a temptation (often succumbed to) to then go ahead and define *intelligence*, but that does not immediately give a clearly grounded meaning to the field. In fact there is danger of deep philosophical regress with no recovery. Therefore I prefer to stay with a more informal notion of intelligence being the sort of stuff that humans do, pretty much all the time.

1.1 Approaches

Traditional Artificial Intelligence has tried to tackle the problem of building artificially intelligent systems from the top down. It tackled intelligence through the notions of *thought* and *reason*. These are things we only know about through introspection. The field has adopted a certain *modus operandi* over the years, which includes a particular set of conventions on how the inputs and outputs to thought and reasoning are to be handled (e.g., the subfield of knowledge representation), and the sorts of things that thought and reasoning do (e.g., planning, problem solving, etc.). I will argue that these conventions cannot account for large aspects of what goes into intelligence. Furthermore, without those aspects the validity of the traditional Artificial Intelligence approaches comes into question. I will also argue that much of the landmark work on thought has been influenced by the technological constraints of the available computers, and thereafter these consequences have often mistakenly become enshrined as principles, long after the original impetus has disappeared.

From an evolutionary stance, human level intelligence did not suddenly leap onto the scene. There were precursors and foundations throughout the lineage to humans. Much of this substrate is present in other animals today. The study of that substrate may well provide constraints on how higher level *thought* in humans could be organized.

Recently there has been a movement to study intelligence from the bottom up, concentrating on physical systems (e.g., mobile robots), situated in the world, autonomously carrying out tasks of various sorts. Some of this work is based on engineering from first principles, other parts of the work are firmly based on biological inspirations. The flavor of this work is quite different from that of traditional Artificial Intelligence. In fact it suggests that despite our best introspections, traditional Artificial Intelligence offers solutions to intelligence which bear almost no resemblance at all to how biological systems work.

There are of course dangers in studying biological systems too closely. Their design was not highly optimized from a global systems point of view. Rather they were patched together and adapted from previously working systems, in ways which most expeditiously met the latest environmental pressures. Perhaps the solutions found for much of intelligence are terribly suboptimal. Certainly there are many vestigial structures surviving within humans' and other animals' digestive, skeletal, and muscular systems. One should suppose then that there are many vestigial neurological structures, interactions, and side effects. Their emulation may be a distraction.

1.2 Outline

The body of this paper is formed by five main sections: 2 *Robots*, 3 *Computers*, 4 *Biology*, 5 *Ideas* and 6 *Thought*. The theme of the paper is how computers and thought have be intimately intertwined in the development of Artificial Intelligence, how those connections may have led the field astray, how biological examples of intelligence are quite different from the models used by Artificial Intelligence, and how recent new approaches point to another path for both computers and thought.

The new approaches that have been developed recently for Artificial Intelligence arose out of work with mobile robots. Section 2 (Robots) briefly outlines the context within which this work arose, and discusses some key realizations made by the researchers involved.

Section 3 (Computers) traces the development of the foundational ideas for Artificial Intelligence, and how they were intimately linked to the technology available for computation. Neither situatedness nor embodiment were easy to include on the original agenda, although their importance was recognized by many early researchers. The early framework with its emphasis on search has remained dominant, and has led to solutions that seem important within the closed world of Artificial Intelligence, but which perhaps are not very relevant to practical applications. The field of Cybernetics with a heritage of very different tools from the early digital computer, provides an interesting counterpoint, confirming the hypothesis that models of thought are intimately tied to the available models of computation.

Section 4 (Biology) is a brief overview of recent developments in the understanding of biological intelligence. It covers material from ethology, psychology, and neuroscience. Of necessity it is not comprehensive, but it is sufficient to demonstrate that the intelligence of biological systems is organized in ways quite different from traditional views of Artificial Intelligence.

Section 5 (Ideas) introduces the two cornerstones to the new approach to Artificial Intelligence, *situatedness* and *embodiment*, and discusses both intelligence and emergence in these contexts.

The last major section, 6 (Thought), outlines some details of the approach of my group at MIT to building complete situated, embodied, artificially intelligent robots. This approach shares much more heritage with biological systems than with what is usually called Artificial Intelligence.

2 Robots

There has been a scattering of work with mobile robots within the Artificial Intelligence community over the years. Shakey from the late sixties at SRI (see [**Nilsson 84**] for a collection of original reports) is perhaps the best known, but other significant efforts include the CART ([**Moravec 82**]) at Stanford and Hilare ([**Giralt, Chatila and Vaisset 84**]) in Toulouse.

All these systems used offboard computers (and thus they could be the largest most powerful computers available at the time and place), and all operated in mostly[1] static environments. All of these robots operated in environments that at least to some degree had been specially engineered for them. They all sensed the world and tried to build two or three dimensional world models of it. Then, in each case, a planner could ignore the actual world, and operate in the model to produce a plan of action for the robot to achieve whatever goal it had been given. In all three of these robots, the generated plans included at least a nominal path through the world model along which it was intended that the robot should move.

Despite the simplifications (static, engineered environments, and the most powerful available computers) all these robots operated excruciatingly slowly. Much of the processing time was consumed in the perceptual end of the systems and in building the world models. Relatively little computation was used in planning and acting.

An important effect of this work was to provide a framework within which other researchers could operate without testing their ideas on real robots, and even without having any access to real robot data. We will call this framework, the *sense-model-plan-act* framework, or *SMPA* for short. See section 3.6 for more details of how the SMPA framework influenced the manner in which robots were built over the following years, and how those robots in turn imposed restrictions on the ways in which intelligent control programs could be built for them.

There was at least an implicit assumption in this early work with mobile robots, that once the simpler case of operating in a static environment had been solved, then the more difficult case of an actively dynamic environment could be tackled. None of these early SMPA systems were ever extended in this way.

Around 1984, a number of people started to worry about the more general problem of organizing intelligence. There was a requirement that intelligence be reactive to dynamic aspects of the environment, that a mobile robot operate on time scales similar to those of animals and humans, and that intelligence be able to generate robust behavior in the face of uncertain sensors, an unpredicted environment, and a changing world. Some of the key realizations about the organization of intelligence were as follows:

- Most of what people do in their day to day lives is not problem-solving or planning, but rather it is routine activity in a relatively benign, but certainly dynamic, world. Furthermore the representations an agent uses of objects in the world need not rely on a semantic correspondence with symbols that the agent possesses, but rather can be defined through interactions of the agent with the world. Agents based on these ideas have achieved interesting performance levels and were built from combinatorial circuits plus a little timing circuitry ([**Agre and Chapman 87**], [**Agre and Chapman 90**]).
- An observer can legitimately talk about an agent's beliefs and goals, even though the agent need not manipulate symbolic data structures at run time. A formal grounding in semantics used for the agent's design can be compiled away. Agents based on these ideas have achieved interesting performance levels and were built from combinatorial circuits plus a little timing circuitry ([**Rosenschein and Kaelbling 86**], [**Kaelbling and Rosenschein 90**]).
- In order to really test ideas of intelligence it is important to build complete agents which operate in dynamic environments using real sensors. Internal world models which are complete representations of the external environment, besides being impossible to obtain, are not at all necessary for agents to act in a competent manner. Many of the actions of an agent are quite separable—coherent intelligence can emerge from subcomponents interacting in the world. Agents based on these ideas have achieved interesting performance levels and were built from combinatorial circuits plus a little timing circuitry ([**Brooks 86**], [**Brooks 90b**], [**Brooks 91a**]).

A large number of others have also contributed to this approach to organizing intelligence. [**Maes 90a**] is the most representative collection.

There is no generally accepted term to describe this style of work. It has sometimes been characterized by the oxymoron *reactive planning*. I have variously used *Robot Beings* [**Brooks and Flynn 89**] and *Artificial Creatures* [**Brooks 90b**]. Related work on non-mobile, but nevertheless active, systems has been called *active vision*, or *animate vision* [**Ballard 89**]. Some workers refer to their beings, or creatures, as *agents*; unfortunately that term is also used by others to refer to somewhat independent components of intelligence within a single physical creature (e.g., the agencies of [**Minsky 86**]). Sometimes the approach is called *behavior-based* as the computational components tend to be direct behavior producing modules[2]. For the remainder of this paper,

[1] In the case of Shakey, experiments included the existence of a gremlin who would secretly come and alter the environment by moving a block to a different location. However, this would usually happen only once, say, in a many hour run, and the robot would not perceive the dynamic act, but rather might later notice a changed world if the change was directly relevant to the particular subtask it was executing. In the case of the CART, the only dynamic aspect of the world was the change in sun angle over long time periods, and this in fact caused the robot to fail as its position estimation scheme was confused by the moving shadows.

[2] Unfortunately this clashes a little with the meaning of

we will simply call the entities of discussion 'robots' or 'behavior-based robots'.

There are a number of key aspects characterizing this style of work.

- **[Situatedness]** The robots are situated in the world—they do not deal with abstract descriptions, but with the here and now of the world directly influencing the behavior of the system.
- **[Embodiment]** The robots have bodies and experience the world directly—their actions are part of a dynamic with the world and have immediate feedback on their own sensations.
- **[Intelligence]** They are observed to be intelligent—but the source of intelligence is not limited to just the computational engine. It also comes from the situation in the world, the signal transformations within the sensors, and the physical coupling of the robot with the world.
- **[Emergence]** The intelligence of the system emerges from the system's interactions with the world and from sometimes indirect interactions between its components—it is sometimes hard to point to one event or place within the system and say that is why some external action was manifested.

Recently there has been a trend to try to integrate traditional symbolic reasoning, on top of a purely reactive system, both with real robots (e.g., **[Arkin 90]**, **[Mitchell 90]**,) and in simulation (e.g, **[Firby 89]**). The idea is that the reactive system handles the real-time issues of being embedded in the world, while the deliberative system does the 'hard' stuff traditionally imagined to be handled by an Artificial Intelligence system. I think that these approaches are suffering from the well known 'horizon effect'—they have bought a little better performance in their overall system with the reactive component, but they have simply pushed the limitations of the reasoning system a bit further into the future. I will not be concerned with such systems for the remainder of this paper.

Before examining this work in greater detail, we will turn to the reasons why traditional Artificial Intelligence adopted such a different approach.

3 Computers

In evolution there is a theory **[Gould and Eldredge 77]** of punctuated equilibria, where most of the time there is little change within a species, but at intervals a subpopulation branches off with a short burst of greatly accelerated changes. Likewise, I believe that in Artificial Intelligence research over the last forty or so years, there have been long periods of incremental work within established guidelines, and occasionally a shift in orientation and assumptions causing a new subfield to branch off. The older work usually continues, sometimes remaining strong, and sometimes dying off gradually. This description of the field also fits more general models of science, such as **[Kuhn 70]**.

The point of this section is that all those steady-state bodies of work rely, sometimes implicitly, on certain philosophical and *technological* assumptions. The founders of the bodies of work are quite aware of these assumptions, but over time as new people come into the fields, these assumptions get lost, forgotten, or buried, and the work takes on a life of its own for its own sake.

In this section I am particularly concerned with how the architecture of our computers influences our choice of problems on which to work, our models of thought, and our algorithms, and how the problems on which we work, our models of thought, and our algorithm choice puts pressure on the development of architectures of our computers.

Biological systems run on massively parallel, low speed computation, within an essentially fixed topology network with bounded depth. Almost all Artificial Intelligence research, and indeed almost all modern computation, runs on essentially Von Neumann architectures, with a large, inactive memory which can respond at very high speed over an extremely narrow channel, to a very high speed central processing unit which contains very little state. When connections to sensors and actuators are also considered, the gap between biological systems and our artificial systems widens.

Besides putting architectural constraints on our programs, even our mathematical tools are strongly influenced by our computational architectures. Most algorithmic analysis is based on the RAM model of computation (essentially a Von Neumann model, shown to be polynomially equivalent to a Turing machine, e.g., **[Hartmanis 71]**). Only in recent years have more general models gained prominence, but they have been in the direction of oracles, and other improbable devices for our robot beings.

Are we doomed to work forever within the current architectural constraints?

Over the past few centuries computation technology has progressed from making marks on various surfaces (chiselling, writing, etc.), through a long evolutionary chain of purely mechanical systems, then electromechanical relay based systems, through vacuum tube based devices, followed by an evolutionary chain of silicon-based devices to the current state of the art.

It would be the height of arrogance and foolishness to assume that we are now using the ultimate technology for computation, namely silicon based integrated circuits, just as it would have been foolish (at least in retrospect) to assume in the 16th century that Napier's Bones were the ultimate computing technology **[Williams 83]**. Indeed the end of the exponential increase in computation speed for uni-processors is in sight, forcing somewhat the large amount of research into parallel approaches to more computation for the dollar, and per second. But there are other more radical possibilities for changes in computation infrastructure[3]. These include computation based

behavior as used by ethologists as an observed interaction with the world, rather than as something explicitly generated.

[3] Equally radical changes have occurred in the past, but admittedly they happened well before the current high levels

on optical switching ([**Gibbs 85**], [**Brady 90**]), protein folding, gene expression, non-organic atomic switching.

3.1 Prehistory

During the early 1940's even while the second world war was being waged, and the first electronic computers were being built for cryptanalysis and trajectory calculations, the idea of using computers to carry out intelligent activities was already on people's minds.

Alan Turing, already famous for his work on computability [**Turing 37**] had discussions with Donald Michie, as early as 1943, and others less known to the modern Artificial Intelligence world as early as 1941, about using a computer to play chess. He and others developed the idea of minimaxing a tree of moves, and of static evaluation, and carried out elaborate hand simulations against human opponents. Later (during the period from 1945 to 1950 at least) he and Claude Shannon communicated about these ideas[4]. Although there was already an established field of mathematics concerning a theory of games, pioneered by Von Neumann [**Von Neumann and Morgenstern 44**], chess had such a large space of legal positions, that even though everything about it is deterministic, the theories were not particularly applicable. Only heuristic and operational programs seemed plausible means of attack.

In a paper titled *Intelligent Machinery*, written in 1948[5], but not published until long after his death [**Turing 70**], Turing outlined a more general view of making computers intelligent. In this rather short insightful paper he foresaw many modern developments and techniques. He argued (somewhat whimsically, to the annoyance of his employers [**Hodges 83**]) for at least some fields of intelligence, and his particular example is the learning of languages, that the machine would have to be embodied, and claimed success "seems however to depend rather too much on sense organs and locomotion to be feasible".

Turing argued that it must be possible to build a thinking machine since it was possible to build imitations of "any small part of a man". He made the distinction between producing accurate electrical models of nerves, and replacing them computationally with the available technology of vacuum tube circuits (this follows directly from his earlier paper [**Turing 37**]), and the assumption that the nervous system can be modeled as a computational system. For other parts of the body he suggests that "television cameras, microphones, loudspeakers", etc., could be used to model the rest of the system. "This would be a tremendous undertaking of course." Even so, Turing notes that the so constructed machine "would still have no contact with food, sex, sport and many other things of interest to the human being". Turing concludes that the best domains in which to explore the mechanization of thought are various games, and cryptanalysis, "in that they require little contact with the outside world"[6].

Turing thus carefully considered the question of embodiment, and for technical reasons chose to pursue aspects of intelligence which could be viewed, at least in his opinion, as purely symbolic. Minimax search, augmented with the idea of pursuing chains of capture to quiescence, and clever static evaluation functions (the *Turochamp* system of David Champernowne and Alan Turing[7], [**Shannon 50**]) soon became the dominant approach to the problem. [**Newell, Shaw and Simon 58**] compared all four known implemented chess playing programs of 1958 (with a total combined experience of six games played), including Turochamp, and they all followed this approach.

The basic approach of minimax with a good static evaluation function has not changed to this day. Programs of this ilk compete well with International Grand Masters. The best of them, *Deep Thought* [**Hsu, Anantharaman, Campbell and Nowatzyk 90**], uses special purpose chips for massive search capabilities, along with a skillful evaluation scheme and selective deepening to direct that search better than in previous programs.

Although Turing had conceived of using chess as a vehicle for studying human thought processes, this notion has largely gotten lost along the way (there are of course exceptions, e.g., [**Wilkins 79**] describes a system which substitutes chess knowledge for search in the middle game—usually there are very few static evaluations, and tree search is mainly to confirm or deny the existence of a mate). Instead the driving force has always been performance, and the most successful program of the day has usually relied on technological advances. Brute force tree search has been the dominant method, itself dominated by the amount of bruteness available. This in turn has been a product of clever harnessing of the latest technology available. Over the years, the current 'champion' program has capitalized on the available hardware. *MacHack-6* [**Greenblatt, Eastlake and Crocker 67**] made use of the largest available fast memory (256K 36 bits words—about a megabyte or so, or $45 by today's standards) and a new comprehensive architecture (the PDP-6) largely influenced by Minsky and McCarthy's requirements for Lisp and symbolic programming. *Chess 4.0* and its descendants [**Slate and Atkin 84**] relied on the running on the world's faster available computer. *Belle* [**Condon and Thompson 84**] used a smaller central computer, but had a custom move generator, built from LSI circuits. Deep Thought, mentioned above as the most recent champion, relies on custom VLSI cir-

of installed base of silicon-based computers.

[4] Norbert Wiener also outlines the idea of minimax in the final note of the original edition of [**Wiener 48**]. However he restricts the idea to a depth of two or three plays—one assumes for practical reasons, as he does express the general notion for n plays. See Section 3.3 for more details on the ways in which cybernetic models of thought were restricted by the computational models at hand.

[5] Different sources cite 1947 and 1948 as the time of writing.

[6] Interestingly, Turing did not completely abstract even a chess playing machine away from embodiment, commenting that "its only organs need be 'eyes' capable of distinguishing the various positions on a specially made board, and means for announcing its own moves".

[7] See *Personal Computing* January 1980, pages 80–81, for a description of this hand simulation of a chess machine.

cuits to handle its move generation and tree search. It is clear that the success and progress in chess playing programs has been driven by technology enabling large tree searches. Few would argue that today's chess programs/hardware systems are very good models for general human thought processes.

There were some misgivings along the way, however. In an early paper **[Selfridge 56]** argues that better static evaluation is the key to playing chess, so that look-ahead can be limited to a single move except in situations close to mate (and one assumes he would include situations where there is capture, and perhaps exchanges, involved). But, he claims that humans come to chess with a significant advantage over computers (the thrust of the paper is on learning, and in this instance on learning to play chess) as they have concepts such as 'value', 'double threat', the 'centre' etc., already formed. Chess to Selfridge is not a disembodied exercise, but one where successful play is built upon a richness of experience in other, perhaps simpler, situations.

There is an interesting counterpoint to the history of computer chess; the game of Go. The search tree for Go is much much larger than for chess, and a good static evaluation function is much harder to define. Go has never worked out well as a vehicle for research in computer game playing—any reasonable crack at it is much more likely to require techniques much closer to those of human thought—mere computer technology advances are not going to bring the minimax approach close to success in this domain (see **[Campbell 83]** for a brief overview).

Before leaving Turing entirely there is one other rather significant contribution he made to the field which in a sense he predated. In **[Turing 50]** poses the question "Can machines think?". To tease out an acceptable meaning for this question he presented what has come to be known as the *Turing test*, where a person communicates in English over a teletype with either another person or a computer. The goal is to guess whether it is a person or a computer at the other end. Over time this test has come to be an informal goal of Artificial Intelligence[8]. Notice that it is a totally disembodied view of intelligence, although it is somewhat situated in that the machine has to respond in a timely fashion to its interrogator. Turing suggests that the machine should try to simulate a person by taking extra time and making mistakes with arithmetic problems. This is the version of the Turing test that is bandied around by current day Artificial Intelligence researchers[9].

Turing advances a number of strawman arguments against the case that a digital computer might one day be able to pass this test, but he does not consider the need that the machine be fully embodied. In principle, of course, he is right. But how a machine might be then programmed is a question. Turing provides an argument that programming the machine by hand would be impractical, so he suggests having it learn. At this point he brings up the need to embody the machine in some way. He rejects giving it limbs, but suspects that eyes would be good, although not entirely necessary. At the end of the paper he proposes two possible paths towards his goal of a "thinking" machine. The unembodied path is to concentrate on programming intellectual activities like chess, while the embodied approach is to equip a digital computer "with the best sense organs that money can buy, and then teach it to understand and speak English". Artificial Intelligence followed the former path, and has all but ignored the latter approach[10].

3.2 Establishment

The establishment of Artificial Intelligence as a discipline that is clearly the foundation of today's discipline by that name occurred during the period from the famous 'Dartmouth Conference' of 1956 through the publication of the book "Computers and Thought" in 1963 (**[Feigenbaum and Feldman 63]**).

Named and mostly organized by John McCarthy as "The Dartmouth Summer Research Project on Artificial Intelligence" the six-week long workshop brought together those who would establish and lead the major Artificial Intelligence research centers in North America for the next twenty years. McCarthy jointly established the MIT Artificial Intelligence Laboratory with Marvin Minsky, and then went on to found the Stanford Artificial Intelligence Laboratory. Allen Newell and Herbert Simon shaped and lead the group that turned into the Computer Science department at Carnegie-Mellon University. Even today a large portion of the researchers in Artificial Intelligence in North America had one of these four people on their doctoral committee, or were advised by someone who did. The ideas expressed at the Dartmouth meeting have thus had a signal impact upon the field first named there.

As can be seen from interviews of the participants published in **[McCorduck 79]** there is still some disagreement over the intellectual property that was brought to the conference and its relative significance. The key outcome was the acceptance and rise of search as the pre-eminent tool of Artificial Intelligence. There was a general acceptance of the use of search to solve problems, and with this there was an essential abandonment of any notion of situatedness.

Minsky's earlier work had been involved with neural modeling. His Ph.D. thesis at Princeton was concerned with a model for the brain **[Minsky 54]**. Later, while at Harvard he was strongly influenced by McCulloch and Pitts (see **[McCulloch and Pitts 43]**), but by the time of the Dartmouth meeting he had become more involved with symbolic search-based systems. In his collection **[Minsky 68]** of versions of his students' Ph.D. theses, all were concerned to some degree with defining and controlling an appropriate search space.

---

[8] Turing expresses his own belief that it will be possible for a machine with 10^9 bits of store to pass a five minute version of the test with 70% probability by about the year 2000.

[9] In fact there is a yearly competition with a $100,000 prize for a machine that can pass this version of the Turing test.

[10] An excerpt from Turing's paper is reprinted in **[Hofstadter and Dennett 81]**. They leave out the whole section on learning and embodiment.

Simon and Newell presented their recent work on the *Logic Theorist* [**Newell, Shaw and Simon 57**], a program that proved logic theorems by searching a tree of subgoals. The program made extensive use of heuristics to prune its search space. With this success, the idea of heuristic search soon became dominant within the still tiny Artificial Intelligence community.

McCarthy was not so affected by the conference that he had organized, and continues to this day to concentrate on epistemological issues rather than performance programs. However he was soon to invent the Lisp programming language [**McCarthy 1960**] which became the standard model of computation for Artificial Intelligence. It had great influence on the models of thought that were popular however, as it made certain things such as search, and representations based on individuals, much easier to program.

At the time, most programs were written in assembly language. It was a tedious job to write search procedures, especially recursive procedures in the machine languages of the day, although some people such as [**Samuel 59**] (another Dartmouth participant) were spectacularly successful. Newell and Simon owed much of their success in developing the Logic Theorist and their later General Problem Solver [**Newell, Shaw and Simon 59**], to their use of an interpreted language (IPL-V—see [**Newell, Shaw and Simon 61**]) which supported complex list structures and recursion. Many of their student's projects reported in [**Feigenbaum and Feldman 63**] also used this language.

McCarthy's Lisp was much cleaner and simpler. It made processing lists of information and recursive tree searches trivial to program–often a dozen lines of code could replace many hundreds of lines of assembler code. Search procedures now became even easier and more convenient to include in Artificial Intelligence programs. Lisp also had an influence on the classes of representational systems used, as is described in section 3.5.

In [**Minsky 61**], Artificial Intelligence was broken into five key topics: search, pattern recognition, learning, planning and induction. The second through fourth of these were characterized as ways of controlling search (respectively by better selection of tree expansion operators, by directing search through previous experience, and by replacing a given search with a smaller and more appropriate exploration). Again, most of the serious work in Artificial Intelligence according to this breakdown was concerned with search.

Eventually, after much experimentation [**Michie and Ross 70**], search methods became well understood, formalized, and analyzed [**Knuth and Moore 75**], and became celebrated as the primary method of Artificial Intelligence [**Nilsson 71**].

At the end of the era of establishment, in 1963, Minsky generated an exhaustive annotated bibliography ([**Minsky 63**]) of literature "directly concerned with construction of artificial problem-solving systems"[11]. It contains 925 citations, 890 of which are to scientific papers and books, and 35 of which are to collections of such papers. There are two main points of interest here. First, although the title of the bibliography, "A Selected Descriptor-Indexed Bibliography to the Literature on Artificial Intelligence", refers to Artificial Intelligence, in his introduction he refers to the area of concern as "artificial problem-solving systems". Second, and somewhat paradoxically, the scope of the bibliography is much broader than one would expect from an Artificial Intelligence bibliography today. It includes many items on cybernetics, neuroscience, bionics, information and communication theory, and first generation connectionism.

These two contrasting aspects of the bibliography highlight a trend in Artificial Intelligence that continued for the next 25 years. Out of a soup of ideas on how to build intelligent machines the disembodied and non-situated approach of problem-solving search systems emerged as dominant, at least within the community that referred to its own work as Artificial Intelligence.

With hindsight we can step back and look at what happened. Originally search was introduced as a mechanism for solving problems that arguably humans used some search in solving. Chess and logic theorem proving are two examples we have already discussed. In these domains one does not expect instantaneous responses from humans doing the same tasks. They are not tasks that are situated in the world.

One can debate whether even in these tasks it is wise to rely so heavily on search, as bigger problems will have exponentially bad effects on search time—in fact [**Newell, Shaw and Simon 58**] argue just this, but produced a markedly slower chess program because of the complexity of static evaluation and search control. Some, such as [**Samuel 59**] with his checker's playing program, did worry about keeping things on a human timescale. [**Slagle 63**] in his symbolic integration program, was worried about being economically competitive with humans, but as he points out in the last two paragraphs of his paper, the explosive increase in price/performance ratio for computing was able to keep his programs ahead. In general, performance increases in computers were able to feed researchers with a steadily larger search space, enabling them to feel that they were making progress as the years went by. For any given technology level, a long-term freeze would soon show that programs relying on search had very serious problems, especially if there was any desire to situate them in a dynamic world.

In the last paragraph of [**Minsky 61**] he does bring up the possibility of a situated agent, acting as a "thinking aid" to a person. But again he relies on a performance increase in standard computing methods (this time through the introduction of time sharing) to supply the necessary time relevant computations.

In the early days of the formal discipline of Artificial Intelligence, search was adopted as a basic technology. It was easy to program on digital computers. It lead to reasoning systems which are not easy to shoe-horn into situated agents.

[11]It also acted as the combined bibliography for the papers in [**Feigenbaum and Feldman 63**].

3.3 Cybernetics

There was, especially in the forties and fifties, another discipline which could be viewed as having the same goals as we have identified for Artificial Intelligence—the construction of useful intelligent systems and the understanding of human intelligence. This work, known as *Cybernetics*, had a fundamentally different flavor from the today's traditional Artificial Intelligence.

Cybernetics co-evolved with control theory and statistical information theory—e.g., see **[Wiener 48, 61]**. It is the study of the mathematics of machines, not in terms of the functional components of a machine and how they are connected, and not in terms of what an individual machine can do here and now, and but rather in terms of *all* the possible behaviors that an individual machine can produce. There was a strong emphasis on characterizing a machine in terms of its inputs and outputs, and treating it as a *black box* as far as its internal workings were unobservable. The tools of analysis were often differential or integral equations, and these tools inherently limited cybernetics to situations where the boundary conditions were not changing rapidly. In contrast, they often do so in a system situated in a dynamically changing world—that complexity needs to go somewhere; either into discontinuous models or changed boundary conditions.

Cybernetics arose in the context of regulation of machinery and electronic circuits—it is often characterized by the subtitle of Wiener's book as the study of "control and communication in the animal and the machine". The model of computation at the time of its original development was analog. The inputs to and outputs from the machine to be analyzed were usually thought of as almost everywhere continuous functions with reasonable derivatives, and the mechanisms for automated analysis and modeling were usually things that today would be characterized as analog components. As such there was no notion of symbolic search—any search was couched in terms of minimization of a function. There was also much less of a notion of representation as an abstract manipulable entity than was found in the Artificial Intelligence approaches.

Much of the work in Cybernetics really was aimed at understanding animals and intelligence. Animals were modeled as machines, and from those models, it was hoped to glean how the animals changed their behavior through learning, and how that lead to better adaptation to the environment for the whole organism. It was recognized rather early (e.g., **[Ashby 52]** for an explicit statement) that an organism and its environment must be modeled together in order to understand the behavior produced by the organism—this is clearly an expression of situatedness. The tools of feedback analysis were used (**[Ashby 56]**) to concentrate on such issues as stability of the system as the environment was perturbed, and in particular a system's *homeostasis* or ability to keep certain parameters within prescribed ranges, no matter what the uncontrolled variations within the environment.

With regards to embodiment there were some experiments along these lines. Many cybernetic models of organisms were rather abstract demonstrations of homeostasis, but some were concerned with physical robots. **[Walter 50, 51, 53]**[12] describes robots built on cybernetic principles which demonstrated goal-seeking behavior, homeostasis, and learning abilities.

The complexity and abilities of Walter's physically embodied machines rank with the purely imaginary ones in the first half dozen chapters of **[Braitenberg 84]** three decades later.

The limiting factors in these experiments were twofold; (1) the technology of building small self contained robots when the computational elements were miniature (a relative term) vacuum tubes, and (2) the lack of mechanisms for abstractly describing behavior at a level below the complete behavior, so that an implementation could reflect those simpler components. Thus in the first instance the models of thought were limited by technological barriers to implementing those models, and in the second instance, the lack of certain critical components of a model (organization into submodules) restricted the ability to build better technological implementations.

Let us return to Wiener and analyze the ways in which the mechanisms of cybernetics, and the mechanisms of computation were intimately interrelated in deep and self limiting ways.

Wiener was certainly aware of digital machines[13] even in his earlier edition of **[Wiener 48]**. He compared them to analog machines such as the Bush differential analyzer, and declares that the digital (or *numerical*, as he called them) machines are superior for accurate numerical calculations. But in some deep sense Wiener did not see the flexibility of these machines. In an added chapter in **[Wiener 61]** he discussed the problem of building a self reproducing machine, and in the Cybernetic tradition, reduced the problem to modeling the input/output characteristics of a black box, in particular a non-linear transducer. He related methods for approximating observations of this function with a linear combination of basis non-linear transducers, and then showed that the whole problem could be done by summing and multiplying potentials and averaging over time. Rather than turn to a digital computer to do this he stated that there were some interesting possibilities for multiplication devices using piezo-electric effects. We see then the intimate tying together between models of computation,

[12] Much of the book **[Walter 53]** is concerned with early work on electroencephalography and hopes for its role in revealing the workings of the brain—forty years later these hopes do not seem to have been born out.

[13] In the introduction to **[Wiener 48]** he talks about embodying such machines with photoelectric cells, thermometers, strain gauges and motors in the service of mechanical labor. But, in the text of the book he does not make such a connection with models of organisms. Rather he notes that they are intended for many successive runs, with the memory being cleared out between runs and states that "the brain, under normal circumstances, is not the complete analogue of the computing machine but rather the analogue of a single run on such a machine". His models of digital computation and models of thought are too dis-similar to make the connection that we would today.

i.e., analog computation, and models of the essentials of self-reproduction. It is impossible to tease apart cause and effect from this vantage point. The critical point is the way in which the mathematical proposal is tied to a technological implementation as a certification of the validity of the approach[14].

By the mid sixties it was clear that the study of intelligence, even a study arising from the principles of cybernetics, if it was to succeed needed to be more broad-based in its levels of abstraction and tools of analysis. A good example is **[Arbib 64]**[15]. Even so, he still harbors hope that cybernetic methods may turn out to give an understanding of the "overall coordinating and integrating principles" which interrelate the component subsystems of the human nervous system.

3.4 Abstraction

The years immediately following the Dartmouth conference shaped the field of Artificial Intelligence in a way which has not significantly changed. The next few years, in the main, amplified the abstraction away from situatedness, or connectedness to the world[16]. There were a number of demonstrations along the way which seemed to legitimize this abstraction. In this section I review some of those events, and argue that there were fundamental flaws in the conclusions generally drawn.

At MIT **[Roberts 63]** demonstrated a vision program that could match pre-stored models to visual images of blocks and wedges. This program was the forerunner of all modern vision programs, and it was many years before its performance could be matched by others. It took a grey level image of the world, and extracted a cartoon-like line drawing. It was this line drawing that was then fitted, via an inverse perspective transform to the pre-stored models. To those who saw its results this looked like a straightforward and natural way to process images and to build models (based on the prestored library) of the objective reality in front of the camera.

The unfortunate truth however, is that it is extraordinarily difficult to extract reliable line drawings in any sort of realistic cases of images. In Roberts' case the lighting was carefully controlled, the blocks were well painted, and the background was chosen with care. The images of his blocks produced rather complete line drawings with very little clutter where there should, by human observer standards, be no line elements. Today, after almost thirty years of research on bottom-up, top-down, and middle-out line finders, there is still no line finder that gets such clean results on a single natural image. Real world images are not at all the clean things that our personal introspection tells us they are. It is hard to appreciate this without working on an image yourself[17].

The fallout of Roberts' program working on a very controlled set of images was that people thought that the line detection problem was doable and solved. E.g., **[Evans 68]** cites Roberts in his discussion of how input could obtained for his analogy program which compared sets of line drawings of 2-D geometric figures.

During the late sixties and early seventies the Shakey project **[Nilsson 84]** at SRI reaffirmed the premises of abstract Artificial Intelligence. Shakey, mentioned in section 2, was a mobile robot that inhabited a set of specially prepared rooms. It navigated from room to room, trying to satisfy a goal given to it on a teletype. It would, depending on the goal and circumstances, navigate around obstacles consisting of large painted blocks and wedges, push them out of the way, or push them to some desired location.

Shakey had an onboard black and white television camera as its primary sensor. An offboard computer analyzed the images, and merged descriptions of what was seen into an existing first order predicate calculus model of the world. A planning program, STRIPS, operated on those symbolic descriptions of the world to generate a sequence of actions for Shakey. These plans were translated through a series of refinements into calls to atomic actions in fairly tight feedback loops with atomic sensing operations using Shakey's other sensors such as a bump bar and odometry.

Shakey was considered a great success at the time, demonstrating an integrated system involving mobility, perception, representation, planning, execution, and error recovery.

Shakey's success thus reaffirmed the idea of relying completely on internal models of an external objective reality. That is precisely the methodology it followed, and it appeared successful. However, it only worked because of very careful engineering of the environment. Twenty years later, no mobile robot has been demonstrated matching all aspects of Shakey's performance in a more general environment, such as an office environment.

The rooms in which Shakey operated were bare except for the large colored blocks and wedges. This made the class of objects that had to be represented very simple. The walls were of a uniform color, and carefully lighted, with dark rubber baseboards, making clear boundaries with the lighter colored floor. This meant that very simple and robust vision of trihedral corners between two walls and the floor, could be used for relocalizing the robot in order to correct for drift in the robot's odometric measurements. The blocks and wedges were painted different colors on different planar surfaces. This ensured

[14] With hindsight, an even wilder speculation is presented at the end of the later edition. Wiener suggests that the capital substances of genes and viruses may self reproduce through such a spectral analysis of infra-red emissions from the model molecules that then induce self organization into the undifferentiated magma of amino and nucleic acids available to form the new biological material.

[15] Arbib includes an elegant warning against being too committed to models, even mathematical models, which may turn out to be wrong. His statement that the "mere use of formulas gives no magical powers to a theory" is just as timely today as it was then.

[16] One exception was a computer controlled hand built at MIT, **[Ernst 61]**, and connected to the TX-0 computer. The hand was very much situated and embodied, and relied heavily on the external world as a model, rather than using internal representations. This piece of work seems to have gotten lost, for reasons that are not clear to me.

[17] Try it! You'll be amazed at how bad it is.

that it was relatively easy, especially in the good lighting provided, to find edges in the images separating the surfaces, and thus making it easy to identify the shape of the polyhedron. Blocks and wedges were relatively rare in the environment, eliminating problems due to partial obscurations. The objective reality of the environment was thus quite simple, and the mapping to an internal model of that reality was also quite plausible.

Around the same time at MIT a major demonstration was mounted of a robot which could view a scene consisting of stacked blocks, then build a copy of the scene using a robot arm (see **[Winston 72]**—the program was known as the *copy-demo*). The programs to do this were very specific to the blocks world, and would not have worked in the presence of simple curved objects, rough texture on the blocks, or without carefully controlled lighting. Nevertheless it reinforced the idea that a complete three dimensional description of the world could be extracted from a visual image. It legitimized the work of others, such as **[Winograd 72]**, whose programs worked in a make-believe world of blocks—if one program could be built which understood such a world completely and could also manipulate that world, then it was assumed that programs which assumed that abstraction could in fact be connected to the real world without great difficulty. The problem remained of slowness of the programs due to the large search spaces, but as before, faster computers were always just around the corner.

The key problem that I see with all this work (apart from the use of search) is that it relied on the assumption that a complete world model could be built internally and then manipulated. The examples from Roberts, through Shakey and the copy-demo all relied on very simple worlds, and controlled situations. The programs were able to largely ignore unpleasant issues like sensor uncertainty, and were never really stressed because of the carefully controlled perceptual conditions. No computer vision systems can produce world models of this fidelity for anything nearing the complexity of realistic world scenes—even object recognition is an active and difficult research area. There are two responses to this: (1) eventually computer vision will catch up and provide such world models—I don't believe this based on the biological evidence presented below, or (2) complete objective models of reality are unrealistic—and hence the methods of Artificial Intelligence that rely on such models are unrealistic.

With the rise in abstraction it is interesting to note that it was still quite technologically difficult to connect to the real world for most Artificial Intelligence researchers[18]. For instance, **[Barrow and Salter 70]** describe efforts at Edinburgh, a major Artificial Intelligence center, to connect sensing to action, and the results are extraordinarily primitive by today's standards—both MIT and SRI had major engineering efforts in support of their successful activities. **[Moravec 81]** relates a sad tale of frustration from the early seventies of efforts at the Stanford Artificial Intelligence Laboratory to build a simple mobile robot with visual input.

Around the late sixties and early seventies there was a dramatic increase in the availability of computer processing power available to researchers at reasonably well equipped laboratories. Not only was there a large increase in processing speed and physical memory, but time sharing systems became well established. An individual researcher was now able to work continuously and conveniently on a disembodied program designed to exhibit intelligence. However, connections to the real world were not only difficult and overly expensive, but the physical constraints of using them made development of the 'intelligent' parts of the system slower by at least an order of magnitude, and probably two orders, as compared to the new found power of timesharing. The computers clearly had a potential to influence the models of thought used—and certainly that hypothesis is not contradicted by the sort of micro-world work that actually went on.

3.5 Knowledge

By this point in the history of Artificial Intelligence, the trends, assumptions, and approaches had become well established. The last fifteen years have seen the discipline thundering along on inertia more than anything else. Apart from a renewed flirtation with neural models (see section 3.8 below) there has been very little change in the underlying assumptions about the models of thought. This coincides with an era of very little technical innovation in our underlying models of computation.

For the remainder of section 3, I rather briefly review the progress made over the last fifteen years, and show how it relates to the fundamental issues of situatedness and embodiment brought up earlier.

One problem with micro-worlds is that they are somewhat uninteresting. The blocks world was the most popular micro-world and there is very little that can be done in it other than make stacks of blocks. After a flurry of early work where particularly difficult 'problems' or 'puzzles' were discovered and then solved (e.g., **[Sussman 75]**) it became more and more difficult to do something new within that domain.

There were three classes of responses to this impoverished problem space:

- Move to other domains with equally simple semantics, but with more interesting print names than *block-a* etc. It was usually not the intent of the researchers to do this, but many in fact did fall into this trap. **[Winograd and Flores 86]** expose and criticize a number of such dressings up in the chapter on "Understanding Language".
- Build a more complex semantics into the blocks world and work on the new problems which arise. A rather heroic example of this is **[Fahlman 74]** who included balance, multi-shaped blocks, friction, and the like. The problem with this approach

[18] It is still fairly difficult even today. There are very few turnkey systems available for purchase which connect sensors to reasonable computers, and reasonable computers to actuators. The situation does seem to be rapidly improving however—we may well be just about to step over a significant threshold.

is that the solutions to the 'puzzles' become so domain specific that it is hard to see how they might generalize to other domains.

- Move to the wider world. In particular, represent knowledge about the everyday world, and then build problem solvers, learning systems, etc., that operate in this semantically richer world.

The last of these approaches has spawned possibly the largest recognizable subfield of Artificial Intelligence, known as Knowledge Representation. It has its own conferences. It has theoretical and practical camps. Yet, it is totally ungrounded. It concentrates much of its energies on anomalies within formal systems which are never used for any practical tasks.

[Brachman and Levesque 85] is a collection of papers in the area. The knowledge representation systems described receive their input either in symbolic form or as the output of natural language systems. The goal of the papers seems to be to represent 'knowledge' about the world. However it is totally ungrounded. There is very little attempt to use the knowledge (save in the naive physics **[Hayes 85]**, or qualitative physics **[de Kleer and Brown 84]** areas—but note that these areas too are ungrounded). There is an implicit assumption that someday the inputs and outputs will be connected to something which will make use of them (see **[Brooks 91a]** for an earlier criticism of this approach).

In the meantime the work proceeds with very little to steer it, and much of it concerns problems produced by rather simple-minded attempts at representing complex concepts. To take but one example, there have been many pages written on the problem of penguins being birds, even though they cannot fly. The reason that this is a problem is that the knowledge representation systems are built on top of a computational technology that makes convenient the use of very simple individuals (Lisp atoms) and placing links between them. As pointed out in **[Brooks 90b]**, and much earlier in **[Brooks 91a]**, such a simple approach does not work when the system is to be physically grounded through embodiment. It seems pointless to try to patch up a system which in the long run cannot possibly work. **[Dreyfus 81]**[19] provides a useful criticism of this style of work.

Perhaps the pinnacle of the knowledge-is-everything approach can be found in **[Lenat and Feigenbaum 91]** where they discuss the foundations of a 10-year project to encode knowledge having the scope of a simple encyclopedia. It is a totally unsituated, and totally disembodied approach. Everything the system is to know is through hand-entered units of 'knowledge', although there is some hope expressed that later it will be able to learn itself by reading. **[Smith 91]** provides a commentary on this approach, and points out how the early years of the project have been devoted to finding a more primitive level of knowledge than was previously envisioned for grounding the higher levels of knowledge. It is my opinion, and also Smith's, that there is a fundamental problem still and one can expect continued regress until the system has some form of embodiment.

[19] Endorsement of some of Dreyfus' views should not be taken as whole hearted embrace of all his arguments.

3.6 Robotics

Section 2 outlined the early history of mobile robots. There have been some interesting developments over the last ten years as attempts have been made to embody some theories from Artificial Intelligence in mobile robots. In this section I briefly review some of the results.

In the early eighties the Defense Advanced Research Projects Agency (DARPA) in the US, sponsored a major thrust in building an Autonomous Land Vehicle. The initial task for the vehicle was to run along a paved road in daylight using vision as the primary perceptual sense. The first attempts at this problem (e.g., **[Waxman, Le Moigne and Srinivasan 85]**) followed the SMPA methodology. The idea was to build a three-dimensional world model of the road ahead, then plan a path along it, including steering and velocity control annotations. These approaches failed as it was not possible to recover accurate three-dimensional road models from the visual images. Even under fairly strong assumptions about the class of roads being followed the programs would produce ludicrously wrong results.

With the pressure of getting actual demonstrations of the vehicle running on roads, and of having all the processing onboard, radical changes had to made in the approaches taken. Two separate teams came up with similar approaches, **[Turk, Morgenthaler, Gremban, and Marra 88]** at Martin Marietta, the integrating contractor, and **[Thorpe, Hebert, Kanade, and Shafer 88]** at CMU, the main academic participant in the project, both producing vision-based navigation systems. Both systems operated in picture coordinates rather than world coordinates, and both successfully drove vehicles along the roads. Neither system generated three dimensional world models. Rather, both identified road regions in the images and servo-ed the vehicle to stay on the road. The systems can be characterized as reactive, situated and embodied. **[Horswill and Brooks 88]** describe a system of similar vintage which operates an indoor mobile robot under visual navigation. The shift in approach taken on the outdoor vehicle was necessitated by the realities of the technology available, and the need to get things operational.

Despite these lessons there is still a strong bias to following the traditional Artificial Intelligence SMPA approach as can be seen in the work at CMU on the Ambler project. The same team that adopted a reactive approach to the road following problem have reverted to a cumbersome, complex, and slow complete world modeling approach **[Simmons and Krotkov 91]**.

3.7 Vision

Inspired by the work of **[Roberts 63]** and that on Shakey **[Nilsson 84]**, the vision community has been content to work on scene description problems for many years. The implicit intent has been that when the reasoning systems of Artificial Intelligence were ready, the vision systems would be ready to deliver world models as required, and the two could be hooked together to get a situated, or embodied system.

There are numerous problems with this approach, and too little room to treat them adequately within the space constraints of this paper. The fundamental issue is that Artificial Intelligence and Computer Vision have made an assumption that the purpose of vision is to reconstruct the static external world (for dynamic worlds it is just supposed to do it often and quickly) as a three dimensional world model. I do not believe that this is possible with the generality that is usually assumed. Furthermore I do not think it is necessary, nor do I think that it is what human vision does. Section 4 discusses some of these issues a little more.

3.8 Parallelism

Parallel computers are potentially quite different from Von Neumann machines. One might expect then that parallel models of computation would lead to fundamentally different models of thought. The story about parallelism, and the influence of parallel machines on models of thought, and the influence of models of thought on parallel machines has two and a half pieces. The first piece arose around the time of the early cybernetics work, the second piece exploded in the mid-eighties and we have still to see all the casualties. The last half piece has been pressured by the current models of thought to change the model of parallelism.

There was a large flurry of work in the late fifties and sixties involving linear threshold devices, commonly known as perceptrons. The extremes in this work are represented by **[Rosenblatt 62]** and **[Minsky and Papert 69]**. These devices were used in rough analogy to neurons and were to be wired into networks that learned to do some task, rather than having to be programmed. Adjusting the weights on the inputs of these devices was roughly equivalent in the model to adjusting the synaptic weights where axons connect to dendrites in real neurons—this is currently considered as the likely site of most learning within the brain.

The idea was that the network had specially distinguished inputs and outputs. Members of classes of patterns would be presented to the inputs and the outputs would be given a correct classification. The difference between the correct response and the actual response of the network would then be used to update weights on the inputs of individual devices. The key driving force behind the blossoming of this field was the perceptron convergence theorem that showed that a simple parameter adjustment technique would always let a single perceptron learn a discrimination if there existed a set of weights capable of making that discrimination.

To make things more manageable the networks were often structured as layers of devices with connections only between adjacent layers. The directions of the connections were strictly controlled, so that there were no feedback loops in the network and that there was a natural progression from one single layer that would then be the input layer, and one layer would be the output layer. The problem with multi-layer networks was that there was no obvious way to assign the credit or blame over the layers for a correct or incorrect pattern classification.

In the formal analyses that were carried out (e.g., **[Nilsson 65]** and **[Minsky and Papert 69]**) only a single layer of devices which could learn, or be adjusted, were ever considered. **[Nilsson 65]** in the later chapters did consider multi-layer machines, but in each case, all but one layer consisted of static unmodifiable devices. There was very little work on analyzing machines with feedback.

None of these machines was particularly situated, or embodied. They were usually tested on problems set up by the researcher. There were many abuses of the scientific method in these tests—the results were not always as the researchers interpreted them.

After the publication of **[Minsky and Papert 69]**, which contained many negative results on the capabilities of single layer machines, the field seemed to die out for about fifteen years.

Recently there has been a resurgence in the field starting with the publication of **[Rumelhart and McClelland 86]**.

The new approaches were inspired by a new learning algorithm known as *back propagation* (**[Rumelhart, Hinton and Williams 86]**). This algorithm gives a method for assigning credit and blame in fully connected multi-layer machines without feedback loops. The individual devices within the layers have linearly weighted inputs and a differentiable output function, a sigmoid, which closely matches a step function, or threshold function. Thus they are only slight generalizations of the earlier perceptrons, but their continuous and differentiable outputs enable hill climbing to be performed which lets the networks converge eventually to be able to classify inputs appropriately as trained.

Back propagation has a number of problems; it is slow to learn in general, and there is a learning rate which needs to be tuned by hand in most cases. The effect of a low learning rate is that the network might often get stuck in local minima. The effect of a higher learning rate is that the network may never really converge as it will be able to jump out of the correct minimum as well as it can jump out of an incorrect minimum. These problems combine to make back propagation, which is the cornerstone of modern neural network research, inconvenient for use in embodied or situated systems.

In fact, most of the examples in the new wave of neural networks have not been situated or embodied. There are a few counterexamples (e.g., **[Sejnowksi and Rosenberg 87]**, **[Atkeson 89]** and **[Viola 90]**) but in the main they are not based on back propagation. The most successful recent learning techniques for situated, embodied, mobile robots, have not been based on parallel algorithms at all—rather they use a reinforcement learning algorithm such as Q-learning (**[Watkins 89]**) as for example, **[Kaelbling 90]** and **[Mahadevan and Connell 90]**.

One problem for neural networks becoming situated or embodied is that they do not have a simple translation into time varying perception or action pattern systems. They need extensive front and back ends to equip them to interact with the world—all the cited examples above had such features added to them.

Both waves of neural network research have been heralded by predictions of the demise of all other forms of computation. It has not happened in either case. Both times there has been a bandwagon effect where many people have tried to use the mechanisms that have become available to solve many classes of problems, often without regard to whether the problems could even be solved in principle by the methods used. In both cases the enthusiasm for the approach has been largely stimulated by a single piece of technology, first the perceptron training rule, and then the back propagation algorithm.

And now for the last half-piece of the parallel computation story. The primary hope for parallel computation helping Artificial Intelligence has been the Connection Machine developed by **[Hillis 85]**. This is a SIMD machine, and as such might be thought to have limited applicability for general intelligent activities. Hillis, however, made a convincing case that it could be used for many algorithms having to do with knowledge representation, and that it would speed them up, often to be constant time algorithms. The book describing the approach is exciting, and in fact on pages 4 and 5 of **[Hillis 85]** the author promises to break the Von Neumann bottleneck by making all the silicon in a machine actively compute all the time. The argument is presented that most of the silicon in a Von Neumann machine is devoted to memory, and most of that is inactive most of the time. This was a brave new approach, but it has not survived the market place. New models of the connection machine have large local memories (in the order of 64K bits) associated with each one bit processor (there can be up to 64K processors in a single Connection Machine). Once again, most of the silicon is inactive most of the time. Connection machines are used within Artificial Intelligence laboratories mostly for computer vision where there is an obvious mapping from processors and their NEWS network to pixels of standard digital images. Traditional Artificial Intelligence approaches are so tied to their traditional machine architectures that they have been hard to map to this new sort of architecture.

4 Biology

We have our own introspection to tell us how our minds work, and our own observations to tell us how the behavior of other people and of animals works. We have our own partial theories and methods of explanation[20]. Sometimes, when an observation, internal or external, does not fit our pre-conceptions, we are rather ready to dismiss it as something we do not understand, and do not need to understand.

In this section I will skim over a scattering of recent work from ethology, psychology, and neuroscience, in an effort to indicate how deficient our everyday understanding of behavior really is. This is important to realize because traditional Artificial Intelligence has relied at the very least implicitly, and sometimes quite explicitly, on these folk understandings of human and animal behavior. The most common example is the story about getting from Boston to California (or vice-versa), which sets up an analogy between what a person does mentally in order to *Plan* the trip, and the means-ends method of planning. See **[Agre 91]** for a more detailed analysis of the phenomenon.

4.1 Ethology

Ethology, the study of animal behavior, tries to explain the causation, development, survival value, and evolution of behavior patterns within animals. See **[McFarland 85]** for an easy introduction to modern ethology.

Perhaps the most famous ethologist was Niko Tinbergen (closely followed by his co-Nobel winners Konrad Lorenz and Karl von Frisch). His heirarchical view of intelligence, described in **[Tinbergen 51]**, is often quoted by Artificial Intelligence researchers in support of their own hierarchical theories. However, this approach was meant to be a neurobiologically plausible theory, but it was described in the absence any evidence. Tinbergen's model has largely been replaced in modern ethology by theories of motivational competition, disinhibition, and dominant and sub-dominant behaviors.

There is no completely worked out theory of exactly how the decision is made as to which behavioral pattern (e.g., drinking or eating) should be active in an animal. A large number of experiments give evidence of complex internal and external feedback loops in determining an appropriate behavior. **[McFarland 88]** presents a number of such experiments and demonstrates the challenges for the theories. The experimental data has ruled out the earlier hierarchical models of behavior selection, and current theories share many common properties with the behavior-based approach advocated in this paper.

4.2 Psychology

The way in which our brains work is quite hidden from us. We have some introspection, we believe, to some aspects of our thought processes, but there are certainly perceptual and motor areas that we are quite confident we have no access to[21]. To tease out the mechanisms at work we can do at least two sorts of experiments: we can test the brain at limits of its operational envelop to see how it breaks down, and we can study damaged brains and get a glimpse at the operation of previously integrated components. In fact, some of these observations call into question the reliability of any of our own introspections.

There have been many psychophysical experiments to test the limits of human visual perception. We are all aware of so-called *optical illusions* where our visual apparatus seems to break down. The journal *Perception* regularly carries papers which show that what we perceive is not what we see (e.g., **[Ramachandran and Anstis 85]**). For instance in visual images of a jumping leopard whose spots are made to artificially move about, we perceive them all as individually following the

[20] See **[Churchland 86]** for a discussion of folk psychology.

[21] This contrasts with a popular fad in Artificial Intelligence where all reasoning of a system is supposed to be available to a meta-reasoning system, or even introspectively to the system itself.

leopard. The straightforward model of human perception proposed by **[Marr 82]**, and almost universally accepted by Artificial Intelligence vision researchers, does not account for such results. Likewise it is now clear that the color pathway is separate from the intensity pathway in the human visual system, and our color vision is something of an illusion[22]. We are unaware of these deficiencies—most people are not aware that they have a blind spot in each eye the size of the image of the moon—they are totally inaccessible to our consciousness. Even more surprising, our very notion of consciousness is full of inconsistencies—psychophysical experiments show that our experience of the flow of time as we observe things in the world is an illusion, as we can often consciously perceive things in a temporal order inconsistent with the world as constructed by an experimenter (see **[Dennett and Kinsbourne 90]** for an overview).

We turn now to damaged brains to get a glimpse at how things might be organized. This work can better be termed *neuropsychology*. There is a large body of literature on this subject from which we merely pick out just a few instances here. The purpose is to highlight the fact that the approaches taken in traditional Artificial Intelligence are vastly different from the way the human brain is organized.

The common view in Artificial Intelligence, and particularly in the knowledge representation community, is that there is a central storage system which links together the information about concepts, individuals, categories, goals, intentions, desires, and whatever else might be needed by the system. In particular there is a tendency to believe that the knowledge is stored in a way that is independent from the way or circumstances in which it was acquired.

[McCarthy and Warrington 88] (and a series of earlier papers by them and their colleagues) give cause to doubt this seemingly logical organization. They report on a particular individual (identified as TOB), who at an advanced age developed a semantic deficit in knowledge of living things, but retained a reasonable knowledge of inanimate things. By itself, this sounds perfectly plausible—the semantic knowledge might just be stored in a category specific way, and the animate part of the storage has been damaged. But, it happens that TOB is able to access the knowledge when, for example he was shown a picture of a dolphin—he was able to form sentences using the word 'dolphin' and talk about its habitat, its ability to be trained, and its role in the US military. When verbally asked what a dolphin is, however, he thought it was either a fish or a bird. He has no such conflict in knowledge when the subject is a wheelbarrow, say. The authors argue that since the deficit is not complete but shows degradation, the hypothesis that there is a deficit in a particular type of sensory modality access to a particular category subclass in a single database is not valid. Through a series of further observations they argue that they have shown evidence of modality-specific organization of meaning, besides a category specific organization. Thus knowledge may be duplicated in many places, and may by no means be uniformly accessible. There are examples of where the knowledge is shown to be inconsistent. Our normal introspection does not reveal this organization, and would seem to be at odds with these explanations. Below, we call into question our normal introspection.

[Newcombe and Ratcliff 89] present a long discussion of visuospatial disorders in brain damaged patients. Many of these severely tax the model of a person as an integrated rational agent. One simple example they report is finger agnosia, where a patient may be quite impaired in the way he can carry out conscious simple tasks using their fingers, but could still do things such as thread a needle, or play the piano well. This suggests the existence of multiple parallel channels of control, rather than some centralized finger control box, for instance.

[Teitelbaum, Pellis and Pellis 90] summarize work which shows that rat locomotion involves a number of reflexes. Drugs can be used to shut off many reflexes so that a rat will appear to be unable to move. Almost all stimuli have no effect—the rat simply remains with its limbs in whatever configuration the experimenter has arranged them. However certain very specific stimuli can trigger a whole chain of complex motor interactions—e.g., tilting the surface on which the rats feet are resting to the point where the rat starts to slide will cause the rat to leap. There has also been a recent popularization of the work of **[Sacks 74]** which shows similar symptoms, in somewhat less understood detail, for humans. Again, it is hard to explain these results in terms of a centralized will—rather an interpretation of multiple almost independent agencies such as hypothesized by **[Minsky 86]** seems a better explanation.

Perhaps the most remarkable sets of results are from split brain patients. It has become common knowledge that we all possess a left brain and a right brain, but in patients whose *corpus callosum* has been severed they really do become separate operational brains in their own rights **[Gazzaniga and LeDoux 77]**.

Through careful experimentation it is possible to independently communicate with the two brains, visually with both, and verbally with the left. By setting up experiments where one side does not have access to the information possessed by the other side, it is possible to push hard on the introspection mechanisms. It turns out that the ignorant half prefers to fabricate explanations for what is going on, rather than admit ignorance. These are normal people (except their brains are cut in half), and it seems that they sincerely believe the lies they are telling, as a result of confabulations generated during introspection. One must question then the ordinary introspection that goes on when our brains are intact.

What is the point of all this? The traditional Artificial Intelligence model of representation and organization along centralized lines is not how people are built. Traditional Artificial Intelligence methods are certainly not necessary for intelligence then, and so far they have not really been demonstrated to be sufficient in situated, embodied systems. The organization of humans is

[22]See the techniques used in the current trend of 'colorization' of black and white movie classics for a commercial capitalization on our visual deficiencies.

by definition sufficient—it is not known at all whether it will turn out to be necessary. The point is that we cannot make assumptions of necessity under either approach. The best we can expect to do for a while at least, is to show that some approach is sufficient to produce interesting intelligence.

4.3 Neuroscience

The working understanding of the brain among Artificial Intelligence researchers seems to be that it is an electrical machine with electrical inputs and outputs to the sensors and actuators of the body. One can see this assumption made explicit, for example, in the fiction and speculative writing of professional Artificial Intelligence researchers such as **[Dennett 81]** and **[Moravec 88]**. This view, and further reduction, leads to the very simple models of brain used in connectionism (**[Rumelhart and McClelland 86]**).

In fact, however, the brain is embodied with a much more serious coupling. The brain is situated in a soup of hormones, that influences it in the strongest possible ways. It receives messages encoded hormonally, and sends messages so encoded throughout the body. Our electrocentrism, based on our electronic models of computation, has lead us to ignore these aspects in our informal models of neuroscience, but hormones play a strong, almost dominating, role in determination of behavior in both simple (**[Kravitz 88]**) and higher animals (**[Bloom 76]**)[23].

Real biological systems are not rational agents that take inputs, compute logically, and produce outputs. They are a mess of many mechanisms working in various ways, out of which emerges the behavior that we observe and rationalize. We can see this in more detail by looking both at the individual computational level, and at the organizational level of the brain.

We do not really know how computation is done at the lowest levels in the brain. There is debate over whether the neuron is the functional unit of the nervous system, or whether a single neuron can act as a many independent smaller units (**[Cohen and Wu 90]**). However, we do know that signals are propagated along axons and dendrites at very low speeds compared to electronic computers, and that there are significant delays crossing synapses. The usual estimates for the computational speed of neuronal systems are no more than about 1 KiloHertz. This implies that the computations that go on in humans to effect actions in the subsecond range must go through only a very limited number of processing steps—the network cannot be very deep in order to get meaningful results out on the timescales that routinely occur for much of human thought. On the other hand, the networks seem incredibly richly connected, compared to the connection width of either our electronic systems, or our connectionist models. For simple creatures some motor neurons are connected to tens of percent of the other neurons in the animal. For mammals motor neurons are typically connected to 5,000 and some neurons in humans are connected to as many as 90,000 other neurons (**[Churchland 86]**).

For one very simple animal *Caenorhabditis elegans*, a nematode, we have a complete wiring diagram of its nervous system, including its development stages (**[Wood 88]**). In the hermaphrodite there are 302 neurons and 56 support cells out of the animal's total of 959 cells. In the male there are 381 neurons and 92 support cells out of a total of 1031 cells. Even though the anatomy and behavior of this creature are well studied, and the neuronal activity is well probed, the way in which the circuits control the animal's behavior is not understood very well at all.

Given that even a simple animal is not yet understood one cannot expect to gain complete insight into building Artificial Intelligence by looking at the nervous systems of complex animals. We can, however, get insight into aspects of intelligent behavior, and some clues about sensory systems and motor systems.

[Wehner 87] for instance, gives great insight into the way in which evolution has selected for sensor-neurological couplings with the environment which can be very specialized. By choosing the right sensors, animals can often get by with very little neurological processing, in order to extract just the right information about the here and now around them, for the task at hand. Complex world model building is not possible given the sensors' limitations, and not needed when the creature is appropriately situated.

[Cruse 90] and **[Götz and Wenking 73]** give insight into how simple animals work, based on an understanding at a primitive level of their neurological circuits. These sorts of clues can help us as we try to build walking robots–for examples of such computational neuroethology see **[Brooks 89]** and **[Beer 90]**.

These clues can help us build better artificial systems, but by themselves they do not provide us with a full theory.

5 Ideas

Earlier we identified situatedness, embodiment, intelligence, and emergence, with a set of key ideas that have lead to a new style of Artificial Intelligence research which we are calling behavior-based robots. In this section I expound on these four topics in more detail.

5.1 Situatedness

Traditional Artificial Intelligence has adopted a style of research where the agents that are built to test theories in intelligence are essentially problem solvers that work in an symbolic abstracted domain. The symbols may have referents in the minds of the builders of the systems, but there is nothing to ground those referents in any real world. Furthermore, the agents are not situated in a world at all. Rather they are given a problem, and they solve it. Then, they are given another problem and they solve it. They are not participating in a world as would agents in the usual sense.

[23] See **[Bergland 85]** for a history of theories of the brain, and how they were influenced by the current technologies available to provide explanatory power. Unfortunately this book is marred by the author's own lack of understanding of computation which leads him to dismiss electrical activity of the brain as largely irrelevant to the process of thought.

In these systems there is no external world per se, with continuity, surprises, or ongoing history. The programs deal only with a model world, with its own built-in physics. There is a blurring between the knowledge of the agent and the world it is supposed to be operating in—indeed in many Artificial Intelligence systems there is no distinction between the two—the agent has access to direct and perfect perception, and direct and perfect action. When consideration is given to porting such agents or systems to operate in the world, the question arises of what sort of representation they need of the real world. Over the years within traditional Artificial Intelligence, it has become accepted that they will need an objective model of the world with individuated entities, tracked and identified over time—the models of knowledge representation that have been developed expect and require such a one-to-one correspondence between the world and the agent's representation of it.

The early robots such as Shakey and the Cart certainly followed this approach. They built models of the world, planned paths around obstacles, and updated their estimate of where objects were relative to themselves as they moved. We developed a different approach **[Brooks 86]** where a mobile robot used the world as its own model—continuously referring to its sensors rather than to an internal world model. The problems of object class and identity disappeared. The perceptual processing became much simpler. And the performance of the robot was better in comparable tasks than that of the Cart[24], and with much less computation, even allowing for the different sensing modalities.

[Agre 88] and **[Chapman 90]** formalized these ideas in their arguments for *deictic* (or *indexical-functional* in an earlier incarnation) representations. Instead of having representations of individual entities in the world, the system has representations in terms of the relationship of the entities to the robot. These relationships are both spatial and functional. For instance in Pengi **[Agre and Chapman 87]**, rather than refer to *Bee-27* the system refers to *the-bee-that-is-chasing-me-now*. The latter may or may not be the same bee that was chasing the robot two minutes previously—it doesn't matter for the particular tasks in which the robot is engaged.

When this style of representation is used it is possible to build computational systems which trade off computational depth for computational width. The idea is that the computation can be represented by a network of gates, timers, and state elements. The network does not need long paths from inputs (sensors) to outputs (actuators). Any computation that is capable of being done is done in a very short time span. There have been other approaches which address a similar time-bounded computation issue, namely the *bounded rationality* approach **[Russell 89]**. Those approaches try to squeeze a traditional Artificial Intelligence system into a bounded amount of computation. With the new approach we tend to come from the other direction, we start with very little computation and build up the amount, while staying away from the boundary of computation that takes too long. As more computation needs to be added there is a tendency to add it in breadth (thinking of the computation as being represented by a circuit whose depth is the longest path length in gates from input to output) rather than depth.

A situated agent must respond in a timely fashion to its inputs. Modeling the world completely under these conditions can be computationally challenging. But a world in which it is situated also provides some continuity to the agent. That continuity can be relied upon, so that the agent can use its perception of the world instead of an objective world model. The representational primitives that are useful then change quite dramatically from those in traditional Artificial Intelligence.

The key idea from situatedness is:

The world is its own best model.

5.2 Embodiment

There are two reasons that embodiment of intelligent systems is critical. First, only an embodied intelligent agent is fully validated as one that can deal with the real world. Second, only through a physical grounding can any internal symbolic or other system find a place to bottom out, and give 'meaning' to the processing going on within the system.

The physical grounding of a robot in the world forces its designer to deal with all the issues. If the intelligent agent has a body, has sensors, and has actuators, then all the details and issues of being in the world must be faced. It is no longer possible to argue in conference papers, that the simulated perceptual system is realistic, or that problems of uncertainty in action will not be significant. Instead, physical experiments can be done simply and repeatedly. There is no room for cheating[25]. When this is done it is usual to find that many of the problems that seemed significant are not so in the physical system (typically 'puzzle' like situations where symbolic reasoning seemed necessary tend not to arise in embodied systems), and many that seemed non-problems become major hurdles (typically these concern aspects of perception and action)[26].

A deeper problem is "can there be disembodied mind?". Many believe that what is human about us is very directly related to our physical experiences. For instance **[Johnson 87]** argues that a large amount of our language is actually metaphorically related to our physical connections to the world. Our mental 'concepts' are based on physically experienced exemplars. **[Smith 91]** suggests that without physical grounding there can be no halt to the regress within a knowledge based system as it tries to reason about real world knowledge such

[24] The tasks carried out by this first robot, *Allen*, were of a different class than those attempted by Shakey. Shakey could certainly not have carried out the tasks that Allen did.

[25] I mean this in the sense of causing self-delusion, not in the sense of wrong doing with intent.

[26] In fact, there is some room for cheating as the physical environment can be specially simplified for the robot—and in fact it may be very hard in some cases to identify such self delusions. In some research projects it may be necessary to test a particular class of robot activities, and therefore it may be necessary to build a test environment for the robot. There is a fine and difficult to define line to be drawn here.

as that contained in an encyclopedia (e.g., **[Lenat and Feigenbaum 91]**).

Without an ongoing participation and perception of the world there is no meaning for an agent. Everything is random symbols. Arguments might be made that at some level of abstraction even the human mind operates in this solipsist position. However, biological evidence (see section 4) suggests that the human mind's connection to the world is so strong, and many faceted, that these philosophical abstractions may not be correct.

The key idea from embodiment is:

> *The world grounds regress.*

5.3 Intelligence

[Brooks 91a] argues that the sorts of activities we usually think of as demonstrating intelligence in humans have been taking place for only a very small fraction of our evolutionary lineage. Further, I argue that the 'simple' things to do with perception and mobility in a dynamic environment took evolution much longer to perfect, and that all those capabilities are a necessary basis for 'higher-level' intellect.

Therefore, I proposed looking at simpler animals as a bottom-up model for building intelligence. It is soon apparent, when 'reasoning' is stripped away as the prime component of a robot's intellect, that the dynamics of the interaction of the robot and its environment are primary determinants of the structure of its intelligence.

Earlier, **[Simon 69]** had discussed a similar point in terms of an ant walking along the beach. He pointed out that the complexity of the behavior of the ant is more a reflection of the complexity of its environment than its own internal complexity. He speculated that the same may be true of humans, but within two pages of text had reduced studying human behavior to the domain of crypto-arithmetic problems.

It is hard to draw the line at what is intelligence, and what is environmental interaction. In a sense it does not really matter which is which, as all intelligent systems must be situated in some world or other if they are to be useful entities.

The key idea from intelligence is:

> *Intelligence is determined by the dynamics of interaction with the world.*

5.4 Emergence

In discussing where intelligence resides in an Artificial Intelligence program **[Minsky 61]** points out that "there is never any 'heart' in a program" and "we find senseless loops and sequences of trivial operations". It is hard to point at a single component as the seat of intelligence. There is no homunculus. Rather, intelligence emerges from the interaction of the components of the system. The way in which it emerges, however, is quite different for traditional and behavior-based Artificial Intelligence systems.

In traditional Artificial Intelligence the modules that are defined are information processing, or functional. Typically these modules might be a perception module, a planner, a world modeler, a learner, etc. The components directly participate in functions such as perceiving, planning, modeling, learning, etc. Intelligent behavior of the system, such as avoiding obstacles, standing up, controlling gaze, etc., emerges from the interaction of the components.

In behavior-based Artificial Intelligence the modules that are defined are behavior producing. Typically these modules might be an obstacle avoidance behavior, a standing up behavior, a gaze control behavior, etc. The components directly participate in producing behaviors such as avoiding obstacles, standing up, controlling gaze, etc. Intelligent functionality of the system, such as perception, planning, modeling, learning, etc., emerges from the interaction of the components.

Although this dualism between traditional and behavior-based systems looks pretty it is not completely accurate. Traditional systems have hardly ever been really connected to the world, and so the emergence of intelligent behavior is something more of an expectation in most cases, rather than an established phenomenon. Conversely, because of the many behaviors present in a behavior-based system, and their individual dynamics of interaction with the world, it is often hard to say that a particular series of actions was produced by a particular behavior. Sometimes many behaviors are operating simultaneously, or are switching rapidly **[Horswill and Brooks 88]**.

Over the years there has been a lot of work on emergence based on the theme of self-organization (e.g., **[Nicolis and Prigogine 77]**). Within behavior-based robots there is beginning to be work at better characterizing emergent functionality, but it is still in its early stages, e.g., **[Steels 90a]**. He defines it as meaning that a function is achieved "indirectly by the interaction of more primitive components among themselves and with the world".

It is hard to identify the seat of intelligence within any system, as intelligence is produced by the interactions of many components. Intelligence can only be determined by the total behavior of the system and how that behavior appears in relation to the environment.

The key idea from emergence is:

> *Intelligence is in the eye of the observer.*

6 Thought

Since late 1984 I have been building autonomous mobile robots in the 'Mobot Lab' at the MIT Artificial Intelligence Laboratory; **[Brooks 86]** gives the original ideas, and **[Brooks 90b]** contains a recent summary of the capabilities of the robots developed in my laboratory over the years.

My work fits within the framework described above in terms of situatedness, embodiment, intelligence and emergence. In particular I have advocated situatedness, embodiment, and highly reactive architectures with no reasoning systems, no manipulable representations, no symbols, and totally decentralized computation. This different model of computation has lead to radically different models of thought.

I have been accused of overstating the case that the new approach is all that is necessary to build truly intelligent systems. It has even been suggested that as an evangelist I have deliberately overstated my case to pull people towards the correct level of belief, and that really all along, I have known that a hybrid approach is necessary.

That is not what I believe. I think that the new approach can be extended to cover the whole story, both with regards to building intelligent systems and to understanding human intelligence—the two principal goals identified for Artificial Intelligence at the beginning of the paper.

Whether I am right or not is an empirical question. Multiple approaches to Artificial Intelligence will continue to be pursued. At some point we will be able to evaluate which approach has been more successful.

In this section I want to outline the philosophical underpinnings of my work, and discuss why I believe the approach is the one that will in the end will prove dominant.

6.1 Principles

All research goes on within the constraints of certain principles. Sometimes these are explicit, and sometimes they are implicit. In the following paragraphs I outline as explicitly as I can the principles followed.

The first set of principles defines the domain for the work.

- The goal is to study complete integrated intelligent autonomous agents.
- The agents should be embodied as mobile robots, situated in unmodified worlds found around our laboratory[27]. This confronts the embodiment issue. The environments chosen are for convenience, although we strongly resist the temptation to change the environments in any way for the robots.
- The robots should operate equally well when visitors, or cleaners, walk through their workspace, when furniture is rearranged, when lighting or other environmental conditions change, and when their sensors and actuators drift in calibration. This confronts the situatedness issue.
- The robots should operate on timescales commensurate with the time scales used by humans. This too confronts the situatedness issue.

The specific model of computation used was not originally based on biological models. It was one arrived at by continuously refining attempts to program a robot to reactively avoid collisions in a people-populated environment, **[Brooks 86]**. Now, however, in stating the principles used in the model of computation, it is clear that it shares certain properties with models of how neurological systems are arranged. It is important to emphasize that it only shares certain properties. Our model of computation is not intended as a realistic model of how neurological systems work. We call our computation model the *subsumption architecture* and its purpose is to program intelligent, situated, embodied agents.

Our principles of computation are:

- Computation is organized as an asynchronous network of active computational elements (they are *augmented finite state machines*—see **[Brooks 89]** for details[28]), with a fixed topology network of unidirectional connections.
- Messages sent over connections have no implicit semantics—they are small numbers (typically 8 or 16 bits, but on some robots just 1 bit) and their meanings are dependent on the dynamics designed into both the sender and receiver.
- Sensors and actuators are connected to this network, usually through asynchronous two-sided buffers.

These principles lead to certain consequences. In particular:

- The system can certainly have state—it is not at all constrained to be purely reactive.
- Pointers and manipulable data structures are very hard to implement (since the model is Turing equivalent it is of course possible, but hardly within the spirit).
- Any search space must be quite bounded in size, as search nodes cannot be dynamically created and destroyed during the search process.
- There is no implicit separation of data and computation, they are both distributed over the same network of elements.

In considering the biological observations outlined in section 4, certain properties seemed worth incorporating into the way in which robots are programmed within the given model of computation. In all the robots built in the mobot lab, the following principles of organization of intelligence have been observed:

- There is no central model maintained of the world. All data is distributed over many computational elements.
- There is no central locus of control.
- There is no separation into perceptual system, central system, and actuation system. Pieces of the network may perform more than one of these functions. More importantly, there is intimate intertwining of aspects of all three of them.
- The behavioral competence of the system is improved by adding more behavior-specific network to the existing network. We call this process *layering*. This is a simplistic and crude analogy to evolutionary development. As with evolution, at every stage

[27] This constraint has slipped a little recently as we are working on building prototype small legged planetary rovers (**[Angle and Brooks 90]**). We have built a special purpose environment for the robots—a physically simulated lunar surface.

[28] For programming convenience we use a higher level abstraction known as the *Behavior Language*, documented in **[Brooks 90c]**. It compiles down to a network of machines as described above.

of the development the systems are tested—unlike evolution there is a gentle debugging process available. Each of the layers is a behavior-producing piece of network in its own right, although it may implicitly rely on presence of earlier pieces of network.

- There is no hierarchical arrangement—i.e., there is no notion of one process calling on another as a subroutine. Rather the networks are designed so that needed computations will simply be available on the appropriate input line when needed. There is no explicit synchronization between a producer and a consumer of messages. Message reception buffers can be overwritten by new messages before the consumer has looked at the old one. It is not atypical for a message producer to send 10 messages for every one that is examined by the receiver.
- The layers, or behaviors, all run in parallel. There may need to be a conflict resolution mechanism when different behaviors try to give different actuator commands.
- The world is often a good communication medium for processes, or behaviors, within a single robot.

It should be clear that these principles are quite different to the ones we have become accustomed to using as we program Von Neumann machines. It necessarily forces the programmer to use a different style of organization for their programs for intelligence.

There are also always influences on approaches to building thinking machines that lie outside the realm of purely logical or scientific thought. The following, perhaps arbitrary, principles have also had an influence on the organization of intelligence that has been used in Mobot Lab robots:

- A decision was made early on that all computation should be done onboard the robots. This was so that the robots could run tether-free and without any communication link. The idea is to download programs over cables (although in the case of some of our earlier robots the technique was to plug in a newly written erasable ROM) into non-volatile storage on the robots, then switch them on to interact with and be situated in the environment.
- In order to maintain a long term goal of being able to eventually produce very tiny robots ([**Flynn 87**]) the computational model has been restricted so that any specification within that model could be rather easily compiled into a silicon circuit. This has put an additional constraint on designers of agent software, in that they cannot use non-linear numbers of connections between collections of computational elements, as that would lead to severe silicon compilation problems. Note that the general model of computation outlined above is such that a goal of silicon compilation is in general quite realistic.

The point of section 3 was to show how the technology of available computation had a major impact on the shape of the developing field of Artificial Intelligence. Likewise there have been a number of influences on my own work that are technological in nature. These include:

- Given the smallness in overall size of the robots there is a very real limitation on the amount of onboard computation that can be carried, and by an earlier principle all computation must be done onboard. The limiting factor on the amount of portable computation is not weight of the computers directly, but the electrical power that is available to run them. Empirically we have observed that the amount of electrical power available is proportional to the weight of the robot[29].
- Since there are many single chip microprocessors available including EEPROM and RAM, it is becoming more possible to include large numbers of sensors which require interrupt servicing, local calibration, and data massaging. The microprocessors can significantly reduce the overall wiring complexity by servicing a local group of sensors (e.g., all those on a single leg of a robot) *in situ*, and packaging up the data to run over a communication network to the behavior-producing network.

These principles have been used in the programming of a number of behavior-based robots. Below we point out the importance of some of these robot demonstrations in indicating how the subsumption architecture (or one like it in spirit) can be expected to scale up to very intelligent applications. In what follows individual references are given to the most relevant piece of the literature. For a condensed description of what each of the robots is and how they are programmed, the reader should see [**Brooks 90b**]; it also includes a number of robots not mentioned here.

6.2 Reactivity

The earliest demonstration of the subsumption architecture was on the robot *Allen* ([**Brooks 86**]). It was almost entirely reactive, using sonar readings to keep away from people and other moving obstacles, while not colliding with static obstacles. It also had a non-reactive higher level layer that would select a goal to head towards, and then proceed to that location while the lower level reactive layer took care of avoiding obstacles.

The very first subsumption robot thus combined non-reactive capabilities with reactive ones. But the important point is that it used exactly the same sorts of computational mechanism to do both. In looking at the network of the combined layers there was no obvious partition into lower and higher level components based on the type of information flowing on the connections, or the state machines that were the computational elements. To be sure, there was a difference in function between the two layers, but there was no need to introduce any centralization or explicit representations to

[29] Jon Connell, a former member of the Mobot Lab, plotted data from a large number of mobile robots and noted the empirical fact that there is roughly one watt of electrical power available for onboard computation for every pound of overall weight of the robot. We call this *Connell's Law*.

achieve a higher level, or later, process having useful and effective influence over a lower level.

The second robot, *Herbert* ([**Connell 89**]), pushed on the reactive approach. It used a laser scanner to find soda can-like objects visually, infrared proximity sensors to navigate by following walls and going through doorways, a magnetic compass to maintain a global sense of orientation, and a host of sensors on an arm which were sufficient to reliably pick up soda cans. The task for Herbert was to wander around looking for soda cans, pick one up, and bring it back to where Herbert had started from. It was demonstrated reliably finding soda cans in rooms using its laser range finder (some tens of trials), picking up soda cans many times (over 100 instances), reliably navigating (many hours of runs), and in one finale doing all the tasks together to navigate, locate, pickup and return with a soda can[30].

In programming Herbert it was decided that it should maintain no state longer than three seconds, and that there would be no internal communication between behavior generating modules. Each one was connected to sensors on the input side, and a fixed priority arbitration network on the output side. The arbitration network drove the actuators.

In order to carry out its tasks, Herbert, in many instances, had to use the world as its own best model and as a communication medium. E.g., the laser-based soda can object finder drove the robot so that its arm was lined up in front of the soda can. But it did not tell the arm controller that there was now a soda can ready to be picked up. Rather, the arm behaviors monitored the shaft encoders on the wheels, and when they noticed that there was no body motion, initiated motions of the arm, which in turn triggered other behaviors, so that eventually the robot would pick up the soda can.

The advantage of this approach is was that there was no need to set up internal expectations for what was going to happen next; that meant that the control system could both (1) be naturally opportunistic if fortuitous circumstances presented themselves, and (2) it could easily respond to changed circumstances, such as some other object approaching it on a collision course.

As one example of how the arm behaviors cascaded upon one another, consider actually grasping a soda can. The hand had a grasp reflex that operated whenever something broke an infrared beam between the fingers. When the arm located a soda can with its local sensors, it simply drove the hand so that the two fingers lined up on either side of the can. The hand then independently grasped the can. Given this arrangement, it was possible for a human to hand a soda can to the robot. As soon as it was grasped, the arm retracted—it did not matter whether it was a soda can that was intentionally grasped, or one that magically appeared. The same opportunism among behaviors let the arm adapt automatically to a wide variety of cluttered desktops, and still successfully find the soda can.

In order to return to where it came from after picking up a soda can, Herbert used a trick. The navigation routines could carry implement rules such as: *when passing through a door southbound, turn left*. These rules were conditionalized on the separation of the fingers on the hand. When the robot was outbound with no can in its hand, it effectively executed one set of rules. After picking up a can, it would execute a different set. By carefully designing the rules, Herbert was guaranteed, with reasonable reliability, to retrace its path.

The point of Herbert is two-fold.

- It demonstrates complex, apparently goal directed and intentional, behavior in a system which has no long term internal state and no internal communication.
- It is very easy for an observer of a system to attribute more complex internal structure than really exists. Herbert appeared to be doing things like path planning and map building, even though it was not.

6.3 Representation

My earlier paper [**Brooks 91a**] is often criticized for advocating absolutely no representation of the world within a behavior-based robot. This criticism is invalid. I make it clear in the paper that I reject traditional Artificial Intelligence representation schemes (see section 5). I also made it clear that I reject explicit representations of goals within the machine.

There can, however, be representations which are partial models of the world—in fact I mentioned that "individual layers extract only those *aspects* of the world which they find relevant—projections of a representation into a simple subspace" [**Brooks 91a**]. The form these representations take, within the context of the computational model we are using, will depend on the particular task those representations are to be used for. For more general navigation than that demonstrated by Connell it may sometimes[31] need to build and maintain a map.

[**Mataric 90, 91**] introduced *active-constructive representations* to subsumption in a sonar-based robot, *Toto*, which wandered around office environments building a map based on landmarks, and then used that map to get from one location to another. Her representations were totally decentralized and non-manipulable, and there is certainly no central control which build, maintains, or uses the maps. Rather, the map itself is an active structure which does the computations necessary for any path planning the robot needs to do.

Primitive layers of control let Toto wander around following boundaries (such as walls and furniture clutter) in an indoor environment. A layer which detects landmarks, such as flat clear walls, corridors, etc., runs in parallel. It informs the map layer as its detection certainty exceeds a fixed threshold. The map is represented as a graph internally. The nodes of the graph are computational elements (they are identical little subnetworks

[30] The limiting factor on Herbert was the mechanical seating of its chips—its mean time between chip seating failure was no more than 15 minutes.

[31] Note that we are saying only *sometimes*, not *must*—there are many navigation tasks doable by mobile robots which appear intelligent, but which do not require map information at all.

of distinct augmented finite state machines). Free nodes arbitrate and allocate themselves, in a purely local fashion, to represent a new landmark, and set up topological links to physically neighboring nodes (using a limited capacity switching network to keep the total virtual 'wire length' between finite state machines to be linear in the map capacity). These nodes keep track of where the robot is physically, by observing changes in the output of the landmark detector, and comparing that to predictions they have made by local message passing, and by referring to other more primitive (magnetic compass based) coarse position estimation schemes.

When a higher layer wants the robot to go to some known landmark, it merely 'excites', in some particular way the particular place in the map that it wants to go. The excitation (this is an abstraction programmed into the particular finite state machines used here—it is not a primitive—as such there could be many different types of excitation co-existing in the map, if other types of planning are required) is spread through the map following topological links, estimating total path link, and arriving at the *landmark-that-I'm-at-now* node (a deictic representation) with a recommendation of the direction to travel right now to follow the shortest path. As the robot moves so to does its representation of where it is, and at that new node the arriving excitation tells it where to go next. The map thus bears a similarity to the *internalized plans* of **[Payton 90]**, but it represented by the same computational elements that use it—there is no distinction between data and process. Furthermore Mataric's scheme can have multiple simultaneously active goals—the robot will simply head towards the nearest one.

This work demonstrates the following aspects of behavior-based or subsumption systems:

- Such systems can make predictions about what will happen in the world, and have expectations.
- Such systems can make plans—but they are not the same as traditional Artificial Intelligence plans—see **[Agre and Chapman 90]** for an analysis of this issue.
- Such systems can have goals—see **[Maes 90b]** for another way to implement goals within the approach.
- All these things can be done without resorting to central representations.
- All these things can be done without resorting to manipulable representations.
- All these things can be done without resorting to symbolic representations.

6.4 Complexity

Can subsumption-like approaches scale to arbitrarily complex systems? This is a question that cannot be answered affirmatively right now—just as it is totally unfounded to answer the same question affirmatively in the case of traditional symbolic Artificial Intelligence methods. The best one can do is point to precedents and trends.

There are a number of dimensions along which the scaling question can be asked. E.g.,

- Can the approach work well as the environment becomes more complex?
- Can the approach handle larger numbers of sensors and actuators?
- Can the approach work smoothly as more and more layers or behaviors are added?

We answer each of these in turn in the following paragraphs.

The approach taken at the Mobot Lab has been that from day one always test the robot in the most complex environment for which it is ultimately destined. This forces even the simplest levels to handle the most complex environment expected. So for a given robot and intended environment the scaling question is handled by the methodology chosen for implementation. But there is also the question of how complex are the environments that are targeted for with the current generation of robots. Almost all of our robots have been tested and operated in indoor environments with people unrelated to the research wandering through their work area at will. Thus we have a certain degree of confidence that the same basic approach will work in outdoor environments (the sensory processing will have to change for some sensors) with other forms of dynamic action taking place.

The number of sensors and actuators possessed by today's robots are pitiful when compared to the numbers in even simple organisms such as insects. Our first robots had only a handful of identical sonar sensors and two motors. Later a six legged walking robot was built **[Angle 89]**. It had 12 actuators and 20 sensors, and was successfully programmed in subsumption (**[Brooks 89]**) to walk adaptively over rough terrain. The key was to find the right factoring into sensor and actuator subsystems so that interactions between the subsystems could be minimized. A new six legged robot, recently completed (**[Brooks and Angle 90]**, is much more challenging, but still nowhere near the complexity of insects. It has 23 actuators and over 150 sensors. With this level of sensing it is possible to start to develop some of the 'senses' that animals and humans have, such as a kinesthetic sense—this comes from the contributions of many sensor readings. Rather, than feed into a geometric model the sensors feed into a estimate of bodily motion. There is also the question of the types of sensors used. **[Horswill and Brooks 88]** generalized the subsumption architecture so that some of the connections between processing elements could be a *retina bus*, a cable that transmitted partially processed images from one site to another within the system. The robot so programmed was able to follow corridors and follow moving objects in real time.

As we add more layers we find that the interactions can become more complex. **[Maes 89]** introduced the notion of switching whole pieces of the network on and off, using an *activation* scheme for behaviors. That idea is now incorporated into the subsumption methodology **[Brooks 90c]**, and provides a way of implementing both competition and cooperation between behaviors. At a lower level a hormone-like system has been introduced ([**Brooks 91b**]) which models the hormone system of the lobster [**Kravitz 88**] ([**Arkin 88**] had implemented a system with similar inspiration). With these additional control mechanisms we have certainly bought ourselves breathing room to increase the performance of our systems markedly. The key point about these control systems is that they fit exactly into the existing structures, and are totally distributed and local in their operations.

6.5 Learning

Evolution has decided that there is a tradeoff between what we know through our genes and what we must find out for ourselves as we develop. We can expect to see a similar tradeoff for our behavior-based robots.

There are at least four classes of things that can be learned:

1. representations of the world that help in some task
2. aspects of instances of sensors and actuators (this is sometimes called calibration)
3. the ways in which individual behaviors should interact
4. new behavioral modules

The robots in the Mobot Lab have been programmed to demonstrate the first three of these types of learning. The last one has not yet been successfully tackled[32].

Learning representations of the world was already discussed above concerning the work of [**Mataric 90, 91**]. The next step will be to generalize active-constructive representations to more classes of use.

[**Viola 90**] demonstrated calibration of a complex head-eye system modeling the primate vestibulo-ocular system. In this system there is one fast channel between a gyroscope and a high performance pan-tilt head holding the camera, and a slower channel using vision which produces correction signals for the gyroscope channel. The same system was used to learn how to accurately saccade to moving stimuli.

Lastly, [**Maes and Brooks 90**] programmed an early six legged robot to learn to walk using the subsumption architecture along with the behavior activation schemes of [**Maes 89**]. Independent behaviors on each leg monitored the activity of other behaviors and correlated that, their own activity state, and the results from a belly switch which provided negative feedback, as input to a local learning rule which learned under which conditions it was to operate the behavior. After about 20 trials per leg, spread over a total of a minute or two, the robot reliably learns the alternating tripod gait—it slowly seems to emerge out of initially chaotic flailing of the legs.

Learning within subsumption is in its early stages but it has been demonstrated in a number of different critical modes of development.

[32]We did have a failed attempt at this through simulated evolution—this is the approach taken by many in the Artificial Life movement.

6.6 Vistas

The behavior-based approach has been demonstrated on situated embodied systems doing things that traditional Artificial Intelligence would have tackled in quite different ways. What are the key research areas that need to be addressed in order to push behavior-based robots towards more and more sophisticated capabilities?

In this section we outline research challenges in three categories or levels [33]:

- Understanding the dynamics of how an individual behavior couples with the environment via the robot's sensors and actuators. The primary concerns here are what forms of perception are necessary, and what relationships exist between perception, internal state, and action (i.e., how behavior is specified or described).
- Understanding how many behaviors can be integrated into a single robot. The primary concerns here are how independent various perceptions and behaviors can be, how much they must rely on, and interfere with each other, how a competent complete robot can be built in such a way as to accommodate all the required individual behaviors, and to what extent apparently complex behaviors can emerge from simple reflexes.
- Understanding how multiple robots (either a homogeneous, or a heterogeneous group) can interact as they go about their business. The primary concerns here are the relationships between individuals' behaviors, the amount and type of communication between robots, the way the environment reacts to multiple individuals, and the resulting patterns of behavior and their impacts upon the environment (which night not occur in the case of isolated individuals).

Just as research in Artificial Intelligence is broken into subfields, these categories provide subfields of behavior-based robots within which it is possible to concentrate a particular research project. Some of these topics are theoretical in nature, contributing to a science of behavior-based systems. Others are engineering in nature, providing tools and mechanisms for successfully building and programming behavior-based robots. Some of these topics have already been touched upon by researchers in behavior-based approaches, but none of them are yet solved or completely understood.

At the individual behavior level some of the important issues are as follows:

Convergence: Demonstrate or prove that a specified behavior is such that the robot will indeed carry out the desired task successfully. For instance, we may want to give some set of initial conditions for a robot, and some limitations on possible worlds in which it is placed, and show that under those conditions, the robot is guaranteed to follow a particular wall, rather than diverge and get lost.

[33]The reader is referred to [**Brooks 90a**] for a more complete discussion of these issues.

Synthesis: Given a particular task, automatically derive a behavior specification for the creature so that it carries out that task in a way which has clearly demonstrable convergence. I do not expect progress in this topic in the near future.

Complexity: Deal with the complexity of real world environments, and sift out the relevant aspects of received sensations rather than being overwhelmed with a multitude of data.

Learning: Develop methods for the automatic acquisition of new behaviors, and the modification and tuning of existing behaviors.

As multiple behaviors are built into a single robot the following issues need to be addressed:

Coherence: Even though many behaviors may be active at once, or are being actively switched on or off, the robot should still appear to an observer to have coherence of action and goals. It should not be rapidly switching between inconsistent behaviors, nor should two behaviors be active simultaneously, if they interfere with each other to the point that neither operates successfully.

Relevance: The behaviors that are active should be relevant to the situation the robot finds itself in—e.g., it should recharge itself when the batteries are low, not when they are full.

Adequacy: The behavior selection mechanism must operate in such a way that the long term goals that the robot designer has for the robot are met—e.g., a floor cleaning robot should successfully clean the floor in normal circumstances, besides doing all the ancillary tasks that are necessary for it to be successful at that.

Representation: Multiple behaviors might want to share partial representations of the world—in fact the representations of world aspects might generate multiple behaviors when activated appropriately.

Learning: The performance of a robot might be improved by adapting the ways in which behaviors interact, or are activated, as a result of experience.

When many behavior-based robots start to interact there are a whole new host of issues which arise. Many of these same issues would arise if the robots were built using traditional Artificial Intelligence methods, but there has been very little published in these areas.

Emergence: Given a set of behaviors programmed into a set of robots, we would like to be able to predict what the global behavior of the system will be, and as a consequence determine the differential effects of small changes to the individual robots on the global behavior.

Synthesis: As at single behavior level, given a particular task, automatically derive a program for the set of robots so that they carry out the task.

Communication: Performance may be increased by increasing the amount of explicit communication between robots, but the relationship between the amount of communication increase and performance increase needs to be understood.

Cooperation: In some circumstances robots should be able to achieve more by cooperating—the form and specification of such possible cooperations need to be understood.

Interference: Robots may interfere with one another. Protocols for avoiding this when it is undesirable must be included in the design of the creatures' instructions.

Density dependence: The global behavior of the system may be dependent on the density of the creatures and the resources they consume within the world. A characterization of this dependence is desirable. At the two ends of the spectrum it may be the case that (a) a single robot given n units of time performs identically to n robots each given 1 unit of time, and (2) the global task might not be achieved at all if there are fewer than, say, m robots.

Individuality: Robustness can be achieved if all robots are interchangeable. A fixed number of classes of robots, where all robots within a class are identical, is also robust, but somewhat less so. The issue then is to, given a task, decide how many classes of creatures are necessary

Learning: The performance of the robots may increase in two ways through learning. At one level, when one robot learns some skill if might be able to transfer it to another. At another level, the robots might learn cooperative strategies.

These are a first cut at topics of interest within behavior-based approaches. As we explore more we will find more topics, and some that seem interesting now will turn out to be irrelevant.

6.7 Thinking

Can this approach lead to thought? How could it? It seems the antithesis of thought. But we must ask first, what is thought? Like intelligence this is a very slippery concept.

We only know that thought exists in biological systems through our own introspection. At one level we identify thought with the product of our consciousness, but that too is a contentious subject, and one which has had little attention from Artificial Intelligence.

My feeling is that thought and consciousness are epiphenomena of the process of being in the world. As the complexity of the world increases, and the complexity of processing to deal with that world rises, we will see the same evidence of thought and consciousness in our systems as we see in people other than ourselves now. Thought and consciousness will not need to be programmed in. They will emerge.

7 Conclusion

The title of this paper is intentionally ambiguous. The following interpretations all encapsulate important points.

- An earlier paper **[Brooks 91a]**[34] was titled *Intelligence without Representation*. The thesis of that paper was that intelligent behavior could be generated without having explicit manipulable internal representations. *Intelligence without Reason* is thus complementary, stating that intelligent behavior can be generated without having explicit reasoning systems present.
- *Intelligence without Reason* can be read as a statement that intelligence is an emergent property of certain complex systems—it sometimes arises without an easily identifiable reason for arising.
- *Intelligence without Reason* can be viewed as a commentary on the bandwagon effect in research in general, and in particular in the case of Artificial Intelligence research. Many lines of research have become goals of pursuit in their own right, with little recall of the reasons for pursuing those lines. A little grounding occasionally can go a long way towards helping keep things on track.
- *Intelligence without Reason* is also a commentary on the way evolution built intelligence—rather than reason about how to build intelligent systems, it used a generate and test strategy. This is in stark contrast to the way all human endeavors to build intelligent systems must inevitably proceed. Furthermore we must be careful in emulating the results of evolution—there may be many structures and observable properties which are suboptimal or vestigial.

We are a long way from creating Artificial Intelligences that measure up the to the standards of early ambitions for the field. It is a complex endeavor and we sometimes need to step back and question why we are proceeding in the direction we are going, and look around for other promising directions.

Acknowledgements

Maja Mataric reviewed numerous drafts of this paper and gave helpful criticism at every stage of the process. Lynne Parker, Anita Flynn, Ian Horswill and Pattie Maes gave me much constructive feedback on later drafts.

References

[Agre 88] "The Dynamic Structure of Everyday Life", Philip E. Agre, *MIT AI TR-1085*, Oct., 1988.

[Agre 91] "The Dynamic Structure of Everyday Life", Philip E. Agre, *Cambridge University Press*, Cambridge, UK, 1991.

[Agre and Chapman 87] "Pengi: An Implementation of a Theory of Activity", Philip E. Agre and David Chapman, *AAAI-87*, Seattle, WA, 1987, 268–272.

[Agre and Chapman 90] "What Are Plans for?", Philip E. Agre and David Chapman, *in* **[Maes 90a]**, 1990, 17–34.

[34] Despite the publication date it was written in 1986 and 1987, and was complete in its published form in 1987.

[Angle 89] "Genghis, a Six Legged Autonomous Walking Robot", Colin M. Angle, *MIT SB Thesis*, March, 1989.

[Angle and Brooks 90] "Small Planetary Rovers", Colin M. Angle and Rodney A. Brooks, *IEEE/RSJ International Workshop on Intelligent Robots and Systems*, Ikabara, Japan, 1990, 383–388.

[Arbib 64] "Brains, Machines and Mathematics", Michael A. Arbib, *McGraw-Hill*, New York, NY, 1964.

[Arkin 89] "Homeostatic Control for a Mobile Robot: Dynamic Replanning in Hazardous Environments", Ronald C. Arkin, *SPIE Proceedings 1007, Mobile Robots III, William J. Wolfe* (ed), 1989, 407–413.

[Arkin 90] "Integrating Behavioral, Perceptual and World Knowledge in Reactive Navigation", Ronald C. Arkin, *in* **[Maes 90a]**, 1990, 105–122.

[Ashby 52] "Design for a Brain", W. Ross Ashby, *Chapman and Hall*, London, 1952.

[Ashby 56] "An Introduction to Cybernetics", W. Ross Ashby, *Chapman and Hall*, London, 1956.

[Atkeson 89] "Using Local Models to Control Movement", Christopher G. Atkeson, *in Neural Information Processing 2, David S. Touretzky* (ed)*, Morgan Kaufmann*, Los Altos, CA, 1989, 316–324.

[Ballard 89] "Reference Frames for Active Vision", Dana H. Ballard, *Proceedings IJCAI–89*, Detroit, MI, 1989, 1635–1641.

[Barrow and Salter 70] "Design of Low-Cost Equipment for Cognitive Robot Research", H. G. Barrow and S. H. Salter, *Machine Intelligence 5, Bernard Meltzer and Donald Michie* (eds)*, American Elsevier Publishing*, New York, NY, 1970, 555–566.

[Beer 90] "Intelligence as Adaptive Behavior", Randall D. Beer, *Academic Press*, San Diego, CA, 1990.

[Bergland 85] "The Fabric of Mind", Richard Bergland, *Viking*, New York, NY, 1985.

[Bloom 76] "Endorphins: Profound Behavioral Effects", F. E. Bloom, *Science 194*, 1976, 630–634.

[Braitenberg 84] "Vehicles: Experiments in Synthetic Psychology", Valentino Braitenberg, *MIT Press*, Cambridge, MA, 1984.

[Brachman and Levesque 85] "Readings in Knowledge Representation", Ronald J. Brachman and Hector J. Levesque, *Morgan Kaufmann*, Los Altos, CA, 1985.

[Brady 90] "Switching Arrays Make Light Work in a Simple Processor", David Brady, *Nature 344*, 1990, 486–487.

[Brooks 86] "A Robust Layered Control System for a Mobile Robot", Rodney A. Brooks, *IEEE Journal of Robotics and Automation, RA-2*, April, 1986, 14–23.

[Brooks 89] "A Robot that Walks: Emergent Behavior from a Carefully Evolved Network", Rodney A. Brooks, *Neural Computation 1:2*, 1989, 253–262.

[Brooks 90a] "Challenges for Complete Creature Architectures", Rodney A. Brooks, *Proc. First Int. Conf.*

on Simulation of Adaptive Behavior, MIT Press, Cambridge, MA, 1990, 434–443.

[Brooks 90b] "Elephants Don't Play Chess", Rodney A. Brooks, *in* **[Maes 90a]**, 1990, 3–15.

[Brooks 90c] "The Behavior Language; User's Guide", Rodney A. Brooks, *MIT A.I. Lab Memo* 1227, 1990.

[Brooks 91a] "Intelligence Without Representation", Rodney A. Brooks, *Artificial Intelligence, 47*, 1991, 139–160.

[Brooks 91b] "Integrated Systems Based on Behaviors", Rodney A. Brooks, *special issue of SIGART on Integrated Intelligent Systems*, July, 1991.

[Brooks and Flynn 89] "Robot Beings", Rodney A. Brooks and Anita M. Flynn, *IEEE/RSJ International Workshop on Intelligent Robots and Systems*, Tsukuba, Japan, 1989, 2–10.

[Campbell 83] "Go", J. A. Campbell, *in Computer Game-Playing: Theory and Practice, M. A. Bramer* (ed)*, Ellis Horwood*, Chichester, UK, 1983.

[Chapman 90] "Vision, Instruction and Action", David Chapman, *MIT AI TR-1085*, June, 1990.

[Churchland 86] "Neurophilosophy", Patricia Smith Churchland, *MIT Press*, Cambridge, MA, 1986.

[Cohen and Wu 90] "One Neuron, Many Units?", Larry Cohen and Jain-young Wu, *Nature 346*, 1990, 108–109.

[Condon and Thompson 84] "Belle", J. H. Condon and Ken Thompson, *in Chess Skill in Man and Machine, P. W. Frey* (ed)*, Springer-Verlag*, 1984.

[Connell 89] "A Colony Architecture for an Artificial Creature", Jonathan H. Connell, *MIT AI TR-1151*, June, 1989.

[Cruse 90] "What Mechanisms Coordinate Leg Movement in Walking Arthropods?", Holk Cruse, *Trends in Neurosciences 13:1*, 1990, 15–21.

[de Kleer and Brown 84] "A Qualitative Physics Based on Confluences", Johann de Kleer and John Seely Brown, *Artificial Intelligence 24*, 1984, 7–83.

[Dennett 78] "Where Am I?", Daniel C. Dennett, *in* **[Hofstadter and Dennett 81]**, 1981.

[Dennett and Kinsbourne 90] "Time and the Observer: the Where and When of Consciousness in the Brain", Daniel Dennett and Marcel Kinsbourne, *Technical Report, Center for Cognitive Studies, Tufts University*, 1990.

[Dreyfus 81] "From Micro-Worlds to Knowledge Representation: AI at an Impasse", Hubert L. Dreyfus, *in Mind Design, John Haugeland* (ed)*, MIT Press*, Cambridge, MA, 1981, 161–204.

[Ernst 61] "MH-1. A Computer-Operated Mechanical Hand", Heinrich A. Ernst, *MIT Ph.D. Thesis*, Dec, 1961.

[Evans 68] "A Program for the Solution of Geometric-Analogy Intelligence Test Questions", Thomas G. Evans, *in* **[Minsky 68]**, 1968, 271–353.

[Fahlman 74] "A Planning System for Robot Construction Tasks", Scott E. Fahlman, *Artificial Intelligence 5*, 1974, 1–50.

[Feigenbaum and Feldman 63] "Computers and Thought", Edward A. Feigenbaum and Julian Feldman, *McGraw-Hill*, New York, NY, 1963.

[Firby 89] "Adaptive Execution in Dynamic Domains", R. James Firby, *Ph.D. Thesis*, Yale, 1989.

[Flynn 87] "Gnat Robots (And How They Will Change Robotics)", Anita M. Flynn, *IEEE Micro Robots and Teleoperators Workshop*, Hyannis, MA, Nov., 1989.

[Gazzaniga and LeDoux 77] "The Integrated Mind", Michael S. Gazzaniga and Joseph E. LeDoux, *Plenum*, New York, NY, 1977.

[Gibbs 85] "Optical Bistability: Controlling Light with Light", H. M. Gibbs, *Academic Press*, New York, NY, 1985.

[Giralt, Chatila and Vaisset 84] "An Integrated Navigation and Motion Control System for Multisensory Robots", Georges Giralt, Raja Chatila, and Marc Vaisset, *Robotics Research 1, Brady and Paul* (eds)*, MIT Press*, Cambridge, MA, 191–214.

[Götz and Wenking 73] "Visual Control of Locomotion in the Walking Fruitfly *Drosophilia*", Karl Georg Götz and Hans Wenking, *Journal of Computational Physiology 85*, 1973, 235–266.

[Gould and Eldredge 77] "Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered", S. J. Gould and N. Eldredge, *Paleobiology 3*, 1977, 115–151.

[Greenblatt, Eastlake and Crocker 67] "The Greenblatt Chess Program", R. D. Greenblatt, D. E. Eastlake and S. D. Crocker, *Am. Fed. Inf. Proc. Soc. Conference Proceedings, 31*, 1967, 801–810.

[Hartmanis 71] "Computational Complexity of Random Access Stored Program Machines", Juris Hartmanis, *Mathematical Systems Theory 5:3*, 1971, 232–245.

[Hayes 85] "The Second Naive Physics Manifesto", Patrick J. Hayes, *in Formal Theories of the Commonsense World, Jerry R. Hobbs and Robert C. Moore* (eds)*, Ablex*, Norwood, NJ, 1985, 1–36.

[Hillis 85] "The Connection Machine", W. Daniel Hillis, *MIT Press*, Cambridge, MA, 1985.

[Hodges 83] "Alan Turing: The Enigma", Andrew Hodges, *Simon and Schuster*, New York, NY, 1983.

[Hofstadter and Dennett 81] "The Mind's I", Douglas R. Hofstadter and Daniel C. Dennett, *Bantam Books*, New York, NY, 1981.

[Horswill and Brooks 88] "Situated Vision in a Dynamic World: Chasing Objects", Ian D. Horswill and Rodney A. Brooks, *AAAI-88*, St Paul, MN, 1988, 796–800.

[Hsu, Anantharaman, Campbell and Nowatzyk 90] "A Grandmaster Chess Machine", Feng-hsiung Hsu, Thomas Anantharaman, Murray Campbell and Andreas Nowatzyk, *Scientific American, 263(4)*, Oct. 1990, 44–50.

[**Johnson 87**] "The Body in the Mind", Mark Johnson, *University of Chicago Press*, Chicago, IL, 1987.

[**Kaelbling 90**] "Learning in Embedded Systems", Leslie Pack Kaelbling, *Ph.D. Thesis*, Stanford, 1990.

[**Kaelbling and Rosenschein 90**] "Action and Planning in Embedded Agents", Leslie Pack Kaelbling and Stanley J. Rosenschein, *in* [**Maes 90a**], 1990, 35–48.

[**Knuth and Moore 75**] "An Analysis of Alpha-Beta Pruning", Donald E. Knuth and Ronald E. Moore, *Artificial Intelligence 6*, 1975, 293–326.

[**Kravitz 88**] "Hormonal Control of Behavior: Amines and the Biasing of Behavioral Output in Lobsters", Edward A. Kravitz, *Science 241*, Sep. 30, 1988, 1775–1781.

[**Kuhn 70**] "The Structure of Scientific Revolutions", Thomas S. Kuhn, *Second Edition, Enlarged, University of Chicago Press*, Chicago, IL, 1970.

[**Lenat and Feigenbaum 91**] "On the Thresholds of Knowledge", Douglas B. Lenat and Edward A. Feigenbaum, *Artificial Intelligence, 47*, 1991, 185–250.

[**Maes 89**] "The Dynamics of Action Selection", Pattie Maes, *IJCAI-89*, Detroit, MI, 1989, 991–997.

[**Maes 90a**] "Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back", Pattie Maes *(ed), MIT Press*, Cambridge, MA, 1990.

[**Maes 90b**] "Situated Agents Can Have Goals", Pattie Maes, *in* [**Maes 90a**], 1990, 49–70.

[**Maes and Brooks 90**] "Learning to Coordinate Behaviors", Pattie Maes and Rodney A. Brooks, *AAAI-90*, Boston, MA, 1990, 796–802.

[**Mahadevan and Connell 90**] "Automatic Programming of Behavior-based Robots using Reinforcement Learning", Sridhar Mahadevan and Jonathan Connell, *IBM T.J. Watson Research Report*, Dec., 1990.

[**Marr 82**] "Vision", David Marr, *Freeman*, San Francisco, CA, 1982.

[**Mataric 90**] "Navigation with a Rat Brain: A Neurobiologically-Inspired Model for Robot Spatial Representation", Maja J Mataric, *Proc. First Int. Conf. on Simulation of Adaptive Behavior, MIT Press*, Cambridge, MA, 1990, 169–175.

[**Mataric 91**] "Behavioral Synergy Without Explicit Integration", Maja J Mataric, *special issue of SIGART on Integrated Intelligent Systems*, July, 1991.

[**McCarthy 60**] "Recursive Functions of Symbolic Expressions", John McCarthy, *CACM 3*, 1960, 184–195.

[**McCarthy and Warrington 88**] "Evidence for Modality-Specific Systems in the Brain", Rosaleen A. McCarthy and Elizabeth. K. Warrington, *Nature 334*, 1988, 428–430.

[**McCorduck 79**] "Machines Who Think", Pamela McCorduck, *Freeman*, New York, NY, 1979.

[**McCulloch and Pitts 43**] "A Logical Calculus of the Ideas Immanent in Nervous Activity", W. S. McCulloch and W. Pitts, *Bull. of Math. Biophysics 5*, 1943, 115–137.

[**McFarland 85**] "Animal Behavior", David McFarland, *Benjamin/Cummings*, Menlo Park, CA, 1985.

[**McFarland 88**] "Problems of Animal Behavior", David McFarland, *Lognman*, Harlow, UK, 1988.

[**Michie and Ross 70**] "Experiments with the Adaptive Graph Traverser", Donald Michie and Robert Ross, *Machine Intelligence 5, Bernard Meltzer and Donald Michie* (eds)*, American Elsevier Publishing*, New York, NY, 1970, 301–318.

[**Minsky 54**] "Neural Nets and the Brain Model Problem", Marvin Minsky, *unpublished Ph.D. dissertation, Princeton University*, 1954, available from University Microfilms, Ann Arbor, MI.

[**Minsky 61**] "Steps Toward Artificial Intelligence", Marvin Minsky, *Proc. IRE 49*, Jan. 1961, 8–30, also in [**Feigenbaum and Feldman 63**].

[**Minsky 63**] "A Selected Descriptor-Indexed Bibliography to the Literature on Artificial Intelligence", Marvin Minsky, *in* [**Feigenbaum and Feldman 63**], 1963, 453–523.

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

[**Minsky 86**] "The Society of Mind", Marvin Minsky, *Simon and Schuster*, New York, NY, 1986.

[**Minsky and Papert 69**] "Perceptrons", Marvin Minsky and Seymour Papert, *MIT Press*, Cambridge, MA, 1969.

[**Mitchell 90**] "Becoming Increasingly Reactive", Tom M. Mitchell, *AAAI-90*, Boston, MA, 1990, 1051–1058.

[**Moravec 81**] "Robot Rover Visual Navigation", Hans P. Moravec, *UMI Research Press*, Ann Arbor, MI, 1981.

[**Moravec 82**] "The Stanford Cart and the CMU Rover", Hans P. Moravec, *Proceedings of the IEEE, 71(7)*, 1982, 872–884.

[**Moravec 88**] "Mind Children", Hans P. Moravec, *Harvard University Press*, Cambridge, MA, 1988.

[**Newcombe and Ratcliff 89**] "Freda Newcombe and Graham Ratcliff", Disorders of Visupspatial Analysis, *in Handbook of Neuropsychology, Vol 2, Elsevier*, New York, NY, 1989.

[**Newell, Shaw and Simon 57**] "Empirical Explorations with the Logic Theory Machine", Allen Newell, J. C. Shaw, Herbert Simon, *Proc. Western Joint Computer Conference 15*, 1957, 218–329, also in [**Feigenbaum and Feldman 63**].

[**Newell, Shaw and Simon 58**] "Chess Playing Programs and the Problem of Complexity", Allen Newell, J. C. Shaw, Herbert Simon, *IBM Journal of Research and Development 2*, Oct. 1958, 320–335, also in [**Feigenbaum and Feldman 63**].

[**Newell, Shaw and Simon 59**] "A General Problem-Solving Program for a Computer", Allen Newell, J. C. Shaw, Herbert Simon, *Computers and Automation 8(7)*, 1959, 10–16.

[**Newell, Shaw and Simon 61**] "Information Processing Language V Manual", Allen Newell, J. C. Shaw, Herbert Simon, *Prentice-Hall*, Edgewood Cliffs, NJ, 1961.

[Nicolis and Prigogine 77] "Self-Organization in Nonequilibrium Systems", G. Nicolis and I. Prigogine, *Wiley*, New York, NY, 1977.

[Nilsson 65] "Learning Machines", Nils J. Nilsson, *McGraw-Hill*, New York, NY, 1965.

[Nilsson 71] "Problem-Solving Methods in Artificial Intelligence", Nils J. Nilsson, *McGraw-Hill*, New York, NY, 1971.

[Nilsson 84] "Shakey the Robot", Nils J. Nilsson *(ed), SRI A.I. Center Technical Note 323*, April, 1984.

[Payton 90] "Internalized Plans: A Representation for Action Resources", David W. Payton, *in* **[Maes 90a]**, 1990, 89–103.

[Ramachandran and Anstis 85] "Perceptual Organization in Multistable Apparent Motion", Vilayanur S. Ramachandran and Stuart M. Anstis, *Perception 14*, 1985, 135–143.

[Roberts 63] "Machine Perception of Three-Dimensional Solids", Larry G. Roberts, *MIT Lincoln Laboratory, Technical Report No. 315*, May, 1963.

[Rosenblatt 62] "Principles of Neurodynamics", Frank Rosenblatt, *Spartan*, New York, NY, 1962.

[Rosenschein and Kaelbling 86] "The Synthesis of Machines with Provable Epistemic Properties", Stanley J. Rosenschein and Leslie Pack Kaelbling, *Proc. Conf. on Theoretical Aspects of Reasoning about Knowledge, Joseph Halpern* (ed)*, Morgan Kaufmann*, Los Altos, CA, 1986, 83–98.

[Rumelhart, Hinton and Williams 86] "Learning Internal Representations by Error Propagation", D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *in* **[Rumelhart and McClelland 86]**, 1986, 318–364.

[Rumelhart and McClelland 86] "Parallel Distributed Processing", David E. Rumelhart and James L. McClelland, *MIT Press*, Cambridge, MA, 1986.

[Russell 89] "Execution Architectures and Compilation", Stuart J. Russell, *Proceedings IJCAI–89*, Detroit, MI, 1989, 15–20.

[Sacks 74] "Awakenings", Oliver W. Sacks, *Doubleday*, New York, NY, 1974.

[Samuel 59] "Some Studies in Machine Learning Using the Game of Checkers", Arthur L. Samuel, *IBM Journal of Research and Development 3*, July 1959, 211–229, also in **[Feigenbaum and Feldman 63]**.

[Sejnowski and Rosenberg 87] "Parallel Networks that Learn to Pronounce English Text", T. J. Sejnowski and C. R. Rosenberg, *Complex Systems 1*, 145–168.

[Selfridge 56] "Pattern Recognition and Learning", Oliver G. Selfridge, *Proc. Third London Symp. on Information Theory, Colin Cherry* (ed)*, Academic Press*, New York, NY, 1956, 345–353.

[Shannon 50] "A Chess-Playing Machine", Claude E. Shannon, *Scientific American 182(2)*, February, 1950.

[Simon 69] "The Sciences of the Artificial", Herbert A. Simon, *MIT Press*, Cambridge, MA, 1969.

[Simmons and Krotkov 91] "An Integrated Walking System for the Ambler Planetary Rover", Reid Simmons and Eric Krotkov, *Proc. IEEE Robotics and Automation*, Sacramento, CA, 1991, 2086–2091.

[Slagle 63] "A Heuristic Program that Solves Symbolic Integration Problems in Freshman Calculus", James R. Slagle, *in* **[Feigenbaum and Feldman 63]**, 1963, 191–206 (from a 1961 MIT mathematics Ph.D. thesis).

[Slate and Atkin 84] "Chess 4.5–The Northwestern University Chess Program", David J. Slate and Lawrence R. Atkin, *in Chess Skill in Man and Machine, P. W. Frey* (ed)*, Springer-Verlag*, 1984.

[Smith 91] "The Owl and the Electric Encyclopedia", Brian Cantwell Smith, *Artificial Intelligence, 47*, 1991, 251–288.

[Steels 90a] "Towards a Theory of Emergent Functionality", Luc Steels, *Proc. First Int. Conf. on Simulation of Adaptive Behavior, MIT Press*, Cambridge, MA, 1990, 451–461.

[Steels 90b] "Exploiting Analogical Representations", Luc Steels, *in* **[Maes 90a]**, 1990, 71–88.

[Sussman 75] "A Computer Model of Skill Acquisition", Gerald J. Sussman, *Elsevier*, New York, NY, 1975.

[Teitelbaum, Pellis and Pellis 90] "Can Allied Reflexes Promote the Integration of a Robot's Behavior", Philip Teitelbaum, Vivien C. Pellis and Sergio M. Pellis, *Proc. First Int. Conf. on Simulation of Adaptive Behavior, MIT Press*, Cambridge, MA, 1990, 97–104.

[Thorpe, Hebert, Kanade, and Shafer 88] "Vision and Navigation for the Carnegie-Mellon Navlab", Charles Thorpe, Martial Hebert, Takeo Kanade, and Steven A. Shafer, *IEEE Trans. PAMI, 10(3)*, May 1988, 362–373.

[Tinbergen 51] "The Study of Instinct", Niko Tinbergen, *Oxford University Press*, Oxford, UK, 1951.

[Turing 37] "On Computable Numbers with an Application to the Entscheidungsproblem", Alan M. Turing, *Proc. London Math. Soc. 42*, 1937, 230–65.

[Turing 50] "Computing Machinery and Intelligence", Alan M. Turing, *Mind 59*, Oct. 1950, 433–460, also in **[Feigenbaum and Feldman 63]**.

[Turing 70] "Intelligent Machinery", Alan M. Turing, *Machine Intelligence 5, Bernard Meltzer and Donald Michie* (eds)*, American Elsevier Publishing*, New York, NY, 1970, 3–23.

[Turk, Morgenthaler, Gremban, and Marra 88] "VITS–A Vision System for Autonomous Land Vehicle Navigation", Matthew A. Turk, David G. Morgenthaler, Keith D. Gremban, and Martin Marra, *IEEE Trans. PAMI, 10(3)*, May 1988, 342–361.

[Viola 90] "Adaptive Gaze Control", Paul A. Viola, *MIT SM Thesis*, 1990.

[Von Neumann and Morgenstern 44] "Theory of Games and Economic Behavior", J. von Neumann and

O. Morgenstern, *John Wiley and Sons*, New York, NY, 1944.

[Walter 50] "An Imitation of Life", W. Grey Walter, *Scientific American, 182(5)*, May 1950, 42–45.

[Walter 51] "A Machine That Learns", W. Grey Walter, *Scientific American, 185(2)*, August 1951, 60–63.

[Walter 53] "The Living Brain", W. Grey Walter, *Duckworth*, London, 1953, republished by *Penguin*, Harmondsworth, UK, 1961.

[Watkins 89] "Learning from Delayed Rewards", Christopher Watkins, *Ph.D. Thesis*, King's College, Cambridge, 1989.

[Waxman, Le Moigne and Srinivasan 85] "Visual Navigation of Roadways", Allen M. Waxman, Jacqueline Le Moigne and Babu Srinivasan, *Proc. IEEE Robotics and Automation*, St Louis, MO, 1985, 862–867.

[Wehner 87] "'Matched Filters' – Neural Models of the External World", Rüdiger Wehner, *J. comp. Physiol. A 161*, 1987, 511–531.

[Wiener 48] "Cybernetics", Norbert Wiener, *John Wiley and Sons*, New York, NY, 1948.

[Wiener 61] "Cybernetics", Norbert Wiener, *Second Edition, MIT Press*, Cambridge, MA, 1961.

[Wilkins 79] "Using Patterns and Plans to Solve Problems and Control Search", David E. Wilkins, *Stanford AI Memo 329*, July, 1979.

[Williams 83] "From Napier to Lucas", Michael R. Williams, *Annals of the History of Computing*, (5)3, 1983, 279–96.

[Winograd 72] "Understanding Natural Language", Terry Winograd, *Academic Press*, New York, NY, 1972.

[Winograd and Flores 86] "Understanding Computers and Cognition", Terry Winograd and Fernando Flores, *Addison-Wesley*, Reading, MA, 1986.

[Winston 72] "The MIT Robot", Patrick H. Winston, *Machine Intelligence 7, Bernard Meltzer and Donald Michie* (eds)*, John Wiley and Sons*, New York, NY, 1972, 431–463.

[Winston 84] "Artificial Intelligence", Patrick Henry Winston, *Second Edition, Addison-Wesley*, Reading, MA, 1984.

[Wood 88] "The Nematode *Caenorhabditis Elegans*", William B. Wood, *Cold Spring Harbor Laboratory*, Cold Spring Harbor, NY, 1988.