MIND DESIGN

Philosophy
Psychology
Artificial Intelligence

edited by

JOHN HAUGELAND

A Bradford Book

The MIT Press
Cambridge, Massachusetts
London, England

5

Artificial Intelligence Meets Natural Stupidity

DREW McDERMOTT

AS A FIELD, artificial intelligence has always been on the border of respectability, and therefore on the border of crackpottery. Many critics (Dreyfus, 1972; Lighthill, 1973) have urged that we are over the border. We have been very defensive toward this charge, drawing ourselves up with dignity when it is made and folding the cloak of Science about us. On the other hand, in private we have been justifiably proud of our willingness to explore weird ideas, because pursuing them is the only way to make progress.

Unfortunately, the necessity for speculation has combined with the culture of the hacker in computer science (Weizenbaum, 1976) to cripple our self-discipline. In a young field, self-discipline is not necessarily a virtue, but we are not getting any younger. In the past few years, our tolerance of sloppy thinking has led us to repeat many mistakes over and over. If we are to retain any credibility, this should stop.

This paper is an effort to ridicule some of these mistakes. Almost everyone I know should find himself the target at some point or other; if you don't, you are encouraged to write up your own favorite fault. The three described here I suffer from myself. I hope self-ridicule will be a complete catharsis, but I doubt it. Bad tendencies can be very deep-rooted. Remember, though, if we can't criticize ourselves, someone else will save us the trouble.

From: John Haugeland (ed.), Mind Design
(Cambridge, MA: MIT Press, 1981)

Wishful Mnemonics

A major source of simple-mindedness in AI programs is the use of mnemonics like "UNDERSTAND" or "GOAL" to refer to programs and data structures. This practice has been inherited from more traditional programming applications, in which it is liberating and enlightening to be able to refer to program structures by their purposes. Indeed, part of the thrust of the structured programming movement is to program entirely in terms of purposes at one level before implementing them by the most convenient of the (presumably many) alternative lower-level constructs.

However, in AI our programs to a great degree are problems rather than solutions. If a researcher tries to write an "understanding" program, it isn't because he has thought of a better way of implementing this well-understood task, but because he thinks he can come closer to writing the *first* implementation. If he calls the main loop of his program "UNDERSTAND", he is (until proven innocent) merely begging the question. He may mislead a lot of people, most prominently himself, and enrage a lot of others.

What he should do instead is refer to this main loop as "G0034", and see if he can *convince* himself or anyone else that G0034 implements some part of understanding. Or he could give it a name that reveals its intrinsic properties, like NODE-NET-INTERSECTION-FINDER, it being the substance of his theory that finding intersections in networks of nodes constitutes understanding. If Quillian (1969) had called his program the "Teachable Language Node Net Intersection Finder", he would have saved us some reading (except for those of us fanatic about finding the part on teachability).

Many instructive examples of wishful mnemonics by AI researchers come to mind once you see the point. Remember GPS? (Ernst and Newell, 1969.) By now, "GPS" is a colorless term denoting a particularly stupid program to solve puzzles. But it originally meant "General Problem Solver", which caused everybody a lot of needless excitement and distraction. It should have been called LFGNS—"Local-Feature-Guided Network Searcher".

Compare the mnemonics in Planner (Hewitt, 1972) with those in Conniver (Sussman and McDermott, 1972):

Planner	*Conniver*
GOAL	FETCH & TRY-NEXT
CONSEQUENT	IF-NEEDED
ANTECEDENT	IF-ADDED
THEOREM	METHOD
ASSERT	ADD

It is so much harder to write programs using the terms on the right! When you say (GOAL . . .), you can just feel the enormous power at your fingertips. It is, of course, an illusion.

Of course, Conniver has some glaring wishful primitives, too. Calling "multiple data bases" CONTEXTS was dumb. It implies that, say, sentence understanding in context is really easy in this system.

LISP's mnemonics are excellent in this regard (Levin *et al*, 1965). What if atomic symbols had been called "concepts", or CONS had been called ASSOCIATE? As it is, the programmer has no debts to pay to the system. He can build whatever he likes. There are some minor faults; "property lists" are a little risky; but by now the term is sanitized.

Resolution theorists have been pretty good about wishful mnemonics. They thrive on hitherto meaningless words like RESOLVE and PARAMODULATE, which can only have their humble, technical meaning. There are actually quite few pretensions in the resolution literature (Robinson, 1965). Unfortunately, at the top of their intellectual edifice stands the word "deduction". This is very wishful, but not entirely their fault. The logicians who first misused the term (e.g., in the "deduction" theorem) didn't have our problems; pure resolution theorists don't either. Unfortunately, too many AI researchers took them at their word and assumed that deduction, like payroll processing, had been tamed.

Of course, as in many such cases, the only consequence in the long run was that "deduction" changed in meaning, to become something narrow, technical, and not a little sordid.

As AI progresses (at least in terms of money spent), this malady gets worse. We have lived so long with the conviction that robots are possible, even just around the corner, that we can't help hastening their arrival with magic incantations. Winograd (1971) explored some of the complexity of language in sophisticated detail; and now everyone takes "natural-language interfaces" for granted, though none has been written. Charniak (1972) pointed out some approaches to understanding stories, and now the OWL interpreter includes a "story-understanding module". (And, God help us, a top-level "ego loop" (Sunguroff, 1975)).

Some symptoms of this disease are embarrassingly obvious once the epidemic is exposed. We should avoid, for example, labeling any part of our programs an "understander". It is the job of the text accompanying the program to examine carefully how much understanding is present, how it got there, and what its limits are.

But even seemingly harmless mnemonics should be handled gingerly. Let me explore as an example the ubiquitous "IS-A link", which has mesmerized workers in this field for years (Quillian, 1968; Fahlman, 1975; Winograd, 1975). I shall take examples from Fahlman's treatment, but what I say is criticism of calling the thing "IS-A", not his work in particular.

An IS-A link joins two nodes in a "semantic net" (a by-now emasculated misnomer), thus:

```
DOG
 ↑
 |  IS-A
 |
FIDO
```

which is presumably meant to express "Fido is a dog". However, the *intrinsic* description of this link is "indicator-value pair inheritance link." That is, if the piece of network

```
         HAS-AS-PART
DOG ----------------→ TAIL
 ↑
 |  IS-A
 |
FIDO
```

is present, then implicitly, "Fido has [a] tail" is present as well. Here HAS-AS-PART is the indicator, TAIL the value.

Most readers will think it extreme to object to calling this IS-A. Indeed, a self-disciplined researcher will be safe. But many people have fallen into the following IS-A traps:

Often, a programmer will shut his mind to other interpretations of IS-A, or conclude that IS-A is a very simple concept. Then he begins to write nonsensical networks like

```
STATE-OF-MIND        CONCEPT              SIMPLE-THING
      ↑                  \                     /
      | IS-A         IS-A \                   / IS-A
      |                    \                 /
  HAPPINESS
                           SIMPLE-CONCEPT
                                 ↑
                                 | IS-A
                                 |
                               IS-A
```

or suspiciously wishful networks like

```
BINARY-TRANSITIVE-RELATION
            ↑
            | IS-A
            |
      GREATER-THAN
```

This is an illustration of "contagious wishfulness": because one piece of a system is labeled impressively, the things it interacts with inherit grandiosity. A program called "THINK" is likely inexorably to acquire data structures called "THOUGHTS".

A good test for the disciplined programmer is to try using gensyms in key places and see if he still admires his system. For example, if STATE-OF-MIND is renamed G1073, we might have:

```
  G1073
    ↑
    | INHERITS-INDICATORS
    |
HAPPINESS
```

which looks much more dubious.

Concepts borrowed from human language must shake off a lot of surface-structure dust before they become clear. (See the next section of this paper.) "Is" is a complicated word, syntactically obscure. We use it with great facility, but we don't understand it well enough to appeal to it for clarification of anything. If we want to call attention to the "property inheritance" use, why not just *say* INHERITS-INDICATORS? Then, if we wish, we can prove from a completed system that this captures a large part of what "is a" means.

Another error is the temptation to write networks like this:

```
            LIVES-IN                      HAS-AS-PART
AFRICA <------------ ELEPHANT ------------> TUSKS
                        ^
                        | IS-A
                        |
                      RALPH
```

which people do all the time. It is clear to them that Ralph lives in Africa, the same Africa as all the other elephants, but his tusks are his own. But the network doesn't say this. Woods (1975) discusses errors like this in detail.

People reason circularly about concepts like IS-A. Even if originally they were fully aware they were just naming INHERITS-INDICATORS with a short, friendly mnemonic, they later use the mnemonic to conclude things about "is a." For example, it has been proposed that a first cut at representing "Nixon is a Hitler" is:

```
HITLER
  ^
  | IS-A
  |
NIXON
```

It worked for Fido and Dog, didn't it? But we just can't take stuff out of the IS-A concept that we never put in. I find this diagram worse than useless.

Lest this all seem merely amusing, meditate on the fate of those who have tampered with words before. The behaviorists ruined words like "behavior", "response", and, especially, "learning". They now play happily in a dream world, internally consistent but lost to science. And think about this: if "mechanical translation" had been called "word-by-word text manipulation", the people doing it might still be getting government money.

Unnatural Language

In this section I wish to rail against a pervasive sloppiness in our thinking: the tendency to see in natural language a natural source of problems and solutions. Many researchers tend to talk as if an internal knowledge representation ought to be closely related to the "corresponding" English sentences; and that operations on the structure should resemble human conversation or "word problems". Because the fault here is a disregard for logic, it will be hard for my criticism to be logical and clear. Examples will help.

A crucial problem in internal representation is effective naming of entities. Although every entity can be given a primary name of some kind, much information about it will be derived from knowledge about roles it plays. If two persons marry and have children, they play the role of parents in whatever data structure encodes knowledge of the family. Information about them (such as, "parents are older than their children") will be in terms of PARENT-1 and PARENT-2 (or "mother" and "father" if they are of opposite sexes). The naming problem is to ensure that information about PARENT-1 is applied to the primary name G0073 when it is discovered that G0073 shares a family with G0308.

The "natural-language fallacy" appears here in the urge to identify the naming problem with the problem of resolving references in English-language discourse. Although the two problems must at some remote intersection meet, it seems to me to be a waste of time to focus on their similarities. Yet it is hard to avoid the feeling that our ability to understand "the mother" to mean "Maria" is the same as the internal function of "binding" PARENT-1 to G0073. But it can't be.

The uses of reference in discourse are not the same as those of naming in internal representation. A good reason to have differing referential expressions in natural language is to pick out an object to talk about with the least amount of breath. After all, the speaker already knows exactly what he wants to refer to; if he says, "the left arm of the chair" in one place, "the arm" in another, and "it" in a third, it isn't because he thinks of this object in three different ways. But internally, this is exactly the reason for having multiple names. Different canonical data structures with different names for the constituent entities come to be instantiated to refer to the same thing in different ways. The internal user of such a structure must be careful to avoid seeing two things where one is meant.

In discourse, a speaker will introduce a hand and easily refer to "the finger". Frame theorists and other notation-developers find it marvelous that their system practically gives them "the finger" automatically as a piece of the data structure for "hand". As far as I can see, doing this automatically is the worst way of doing it. First, of course, there are four or five fingers, each with its own name, so "the finger" will be ambiguous. Second, a phrase like "the finger" can be used in so many ways that an automatic evaluation to FINGER 109 will be wasteful at best. There are idioms to worry about, as in, "He raised his hand and gave me the finger". (Are we to conclude that the "default finger in the hand frame" is the middle finger?) But even ignoring them, there are many contexts where "the" just doesn't mean what we would like it to. For example, "He removed his glove and I saw the finger was missing". This is like, "The barn burned to the ground five years ago and was completely rebuilt". There are logics in which the same BARN 1051 can have different denotations in different time periods, but do we really want this clumsiness in the heart of our internal representation?

It seems much smarter to put knowledge about translation from natural language to internal representation in the natural language processor, not in the internal representation. I am using "in" loosely; my intent is to condemn an approach that translates language very superficially (using a little syntax and morphology) and hands it to the data base in that form. Instead, the language routine must draw on knowledge about all parts of the sentence in translating "the finger". Its output must be a directly useful internal representation, probably as remote as possible from being "English-like".

These problems stem from a picture of a program constructed of cooperating modules that "talk to" each other. While this may be a reasonable metaphor in some ways, anyone who has actually written such a program knows that "talking" is a very poor model of the communication. Yet many researchers (most extremely Stansfield, 1975, and Hawkinson, 1975) find English to be the ideal notation in which to encode messages. They are aware that message-passing channels are the most frustrating bottleneck through which intelligence must pass, so they wish their way into the solution: let the modules speak in human tongues! Let them use metaphor, allusion, hints, polite requests, pleading, flattery, bribes, and patriotic exhortations to their fellow modules!

It is hard to say where they have gone wronger, in underestimating language or overestimating computer programs. Language is only occasionally a medium of communication of information; even when it is, the ratio of information to packaging is low. The problem of a language speaker is to get the directed attention of an unprepared hearer and slide some information into his mind in a very short time. Since the major time sink is moving his mouth, the language sacrifices everything else to brevity, forcing the hearer to do much quick thinking to compensate. Furthermore, since the speaker doesn't quite know the organization of his hearer's mind, his phrasing of information and packaging must, except for the most stereotyped conversations, be an artwork of suggestiveness and insight.

Communication between computer programs is under completely different constraints. At the current stage of research, it is ridiculous to focus on anything but raw communication of information; we are unable to identify where more devious, Freudian intercourse might occur. Packaging and encoding of the information are usually already done. Ambiguity is avoidable. Even brevity is unimportant (at least for speed), since a huge structure can be transmitted by passing an internal name or pointer to it shared by sender and receiver. Instead, the whole problem is getting the hearer to notice what it has been told. (Not "understand", but "notice". To appeal to understanding at this low level will doom us to tail-chasing failure.) The new structure handed to the receiver should give it "permission" to make progress on its problem. If the sender could give more detailed instructions, it could just execute them itself. Unfortunately, the latitude this leaves the receiver is wasted if it is too "narrow-minded" to see what it has received. (The 1962 paper by Newell on these topics is still the best.)

Everyone who has written a large AI program will know what I am talking about. In this communication effort, the naming problem can be irritating, since the sender must make sure the receiver understands its terms. But there are so many approaches to solving the problem (for example, by passing translation tables around), which are not open to conversing humans, that it recedes quickly into the background. The frustrations lie elsewhere.

Reference is not the only "unnatural language" problem. A related one is the feeble analysis of concepts like "the" and "a" by most AI researchers. There is a natural inclination to let "the" flag a definite description and "a" an existential quantifier (or occasionally a description). Except for *Dick, Jane, and Sally,* and some of Bertrand Russell's work, this approach is not even an approximation.

First the typical noun phrase is not directly translated into the internal representation at all, and does not wind up as an object name. For example, "Despite the peripatetic nature of American students and their families . . . there remain wide gaps and serious misconceptions in our understanding of other peoples and cultures". (*Media and Methods, 11,* No. 2 (1974), 43). Translating this sentence (whose meaning is transparent) is problematic in the extreme. The author means to allude to the fact that Americans travel a lot, as a way of getting around to the claim that they don't travel enough or well enough. Why? We don't know yet why people talk this way. But translation methods that worked on "the big red block" will never succeed on "the . . . nature of American students".

Second, the difference between "the" and "a" is not the difference between "definite" and "indefinite", except vacuously. For example, what is the difference in meaning between

> "Due to the decrease in the American birthrate in the 1960's, our schools are underutilized".
>
> "Due to a decrease in the American birthrate in the 1960's, our schools are underutilized".

In most respects, they "mean" exactly the same thing, since there can have been only one decrease in the birthrate in the 1960's, and each sentence presupposes that it occurred. But in one the author is assuming we know it already; in the other, he is more casual about whether we do or not. We have no theory at all about what difference this difference makes.

It is unfortunate that a logical back seepage has caused people to see words like "the", "a", "all", "or", "and", etc. as being embellished or ambiguous versions of "iota", "$\exists$", "$\forall$", "$\vee$", and "$\wedge$". To cure yourself of this, try examining two pages of a book for ten-year olds, translating the story as you go into an internal representation. (I found Kenny, 1963, pp. 14–15 useful.) If you can do this without difficulty, your case is hopeless.

The obsession with natural language seems to have caused the feeling that the human use of language is a royal road to the cognitive psyche. I find this analogous to preoccupation with imagery as a way of studying vision. Most AI researchers react with amusement to proposals to explain vision in terms of stored images, reducing the physical eye to the mind's eye. But many of the same people notice themselves talking to themselves in English, and conclude that English is very close to the language of thought.

Clearly, there must be some other notation, different in principle from natural language, or we will have done for the ear what imagery theory does for the eye. No matter how fascinating the structure of consciousness is, it is dangerous to gaze too long into its depths. The puzzles we find there can be solved only by sneaking up on them from behind. As of now, we have no idea at all why people experience their thoughts the way they do, in pictures and words. It will probably turn out to be quite different, even simpler, than what we think now, once we understand why and how people experience their thoughts at all.

In the meantime, for many people, natural language has become the preferred means of stating problems for programs to solve. For example, research that began as a study of visual recognition becomes a study of how people come up with an animal that is white, has hooves, and has one horn in the middle of its head. People can do this (and get "unicorn"), but the fact that they can obviously has nothing to do with visual recognition. In visual recognition, the main problems are guessing that you're looking at an animal in the first place, deciding that thing is a horn and that it belongs to the head, deciding whether to look for hooves, etc. The problem as stated in natural language is just not the same. (For example, the difficulties raised by the fact that I omitted presence or absence of wings from my description are different from the corresponding visual problems.)

Linguists have, I think, suffered from this self-misdirection for years. The standard experimental tool of modern linguistics is the eliciting of judgments of grammaticality from native speakers. Although anyone can learn how to make such judgments fairly quickly, it is plainly not a skill that has anything to do with ability to speak English. The real parser in your head is not supposed to report on its inputs' degree of grammaticality; indeed, normally it doesn't "report" at all in a way accessible to verbalization. It just tries to aid understanding of what it hears as best it can. So the grammaticality judgment task is completely artificial. It doesn't correspond to something people normally do.

Linguists, of course, have a place in their ontology for these judgments. They are a direct road to the seat of linguistic "competence". AI people find this notion dubious. They would be just as suspicious if someone claimed that a good way to measure "visual recognition competence" was to measure the ability of a subject to guess where the cubes were in a scene presented to him as an English description of intensity contours. ("A big steep one in the corner, impetuous but not overbearing".)

Eventually, though, we all trick ourselves into thinking that the statement of a problem in natural language is natural. One form of this self-delusion that I have had difficulty avoiding is the "information-retrieval fixation". It dates from Winograd's (1971) analysis of questions like, "Do I like any pyramids?" as a simple PLANNER program like (THAND (THGOAL (LIKE WINOGRAD ?X)) (THGOAL (IS PYRAMID ?X))). This was entirely justified in the context he was dealing with, but clearly a stopgap. Nonetheless, nowadays, when someone invents a representation or deduction algorithm, he almost always illustrates it with examples like this, couched either in natural language or a simple translation like (THAND . . .).

This tight coupling of internal and external problem statements if taken seriously, reduces the chance of progress on representation and retrieval problems. If a researcher tries to think of his problem as natural-language question answering, he is hurt by the requirement that the answers be the results of straightforward data-base queries. Real discourse is almost never of the literal-minded information-retrieval variety. In real discourse, the context leading up to a question sets the stage for it, and usually affects its meaning considerably. But, since the researcher is not really studying language, he cannot use the natural-language context. The only version of natural language he can have in mind must exclude this example of a conversation between two programmers on a system with six-letter file names:

"Where is the function TRY-NEXT defined?"

"In the file TRYNXT >". (pronounced TRY-NEXT)

"How do you spell 'TRY-NEXT'?"

"Omit the e".

Such contextual and intentional effects are distracting at best for the designer of a data base; presumably they are normally helpful to humans.

The other course is to concentrate on handling the query after it has been translated into (THAND . . .), but if this formula is still thought of as a direct translation of an English question, the approach ignores whatever framework a system might use to focus its computational attention. Generally a program builds or prunes its data structure as it goes, organizing it in such a way that most queries worth making at all can be handled with reasonable efficiency. Just picking the THAND problem out of the blue throws this organization away. This is what happens with the naive natural-language information-retrieval paradigm. A researcher who designs his retrieval algorithm around the case of a completely unmotivated formal query is likely to become preoccupied with problems like the efficient intersection of lists of likable objects and pyramids. (Nevins, 1974, Fahlman, 1975.) In the design of programs whose knowledge is organized around problems, such issues are not nearly so important.

Someone must still work on the context-free English query problem, but there is no reason to expect it to be the same as the data-base retrieval problem. Besides, it might turn out that natural language is not the best notation for information-retrieval requests. Perhaps we should postpone trying to get computers to speak English, and try programming librarians in PL/1!

In this section I have been harsh toward AI's tendency to over-simplify or overglorify natural language, but don't think that my opinion is that research in this area is futile. Indeed, probably because I am an academic verbalizer, I feel that understanding natural language is the most fascinating and important research goal we have in the long run. But it deserves more attention from a theoretical point of view before we rush off and throw together "natural-language" interfaces to programs with inadequate depth. We should do more studies of what language is for, and we should develop complex programs with a need to talk, before we put the two together.

"**Only a Preliminary Version of the Program was Actually Implemented"

A common idiocy in AI research is to suppose that having identified the shortcomings of Version I of a program is equivalent to having written Version II. (McDermott, 1974a, Sussman, 1975, Goldstein, 1974.) Naturally, the sincere researcher doesn't think of his actions this way. From my own experience, the course of a piece of research is like this:

Having identified a problem, the ambitious researcher stumbles one day upon a really good idea that neatly solves several related subproblems of it at once. (Sometimes the solution actually comes before the problem is identified.) The idea is formally pretty and seems to mesh smoothly with the way a rational program ought to think. Let us call it "sidetracking control structure" for concreteness. The researcher immediately implements an elegant program embodying automatic sidetracking, with an eye toward applying it to his original problem. As always, implementation takes much longer than expected, but matters are basically tidy.

However, as he develops and debugs this piece of code, he becomes aware that there are several theoretical holes in his design and that it doesn't work. It doesn't work for good and respectable reasons, most of them depending on the fact that the solution to the problem requires more than one good idea. But, having got a framework, he becomes more and more convinced that those small but numerous holes are where the good ideas are to fit. He may even be right.

Here, however, he begins to lose his grip. Implementing Version I, whose shortcomings are all too obvious, was exhausting; it made him feel grubby for nothing. (Not at all like the TECO macros he took time out for along the way!) He feels as though he's paid his dues; now he can join the theoreticians. What's more, he *should.*

Implementation details will make his thesis dull. The people want *epistemology.*

Simultaneously, he enjoys the contradictory feeling that the implementation of Version II would be easy. He has reams of notes on the holes in Version I and how to fill them. When he surveys them, he feels their master. Though a stranger to the trees, he can talk with confidence about the forest. Indeed, that is precisely what he does in his final document. It is full of allusions to a program he seems to be claiming to have written. Only in a cautious footnote does he say, "the program was never actually finished", or, "a preliminary version of the program was actually written".

This final report can have interesting quirks. It is likely to be titled *A Side-Tracking Control Structure Approach to Pornographic Question-Answering,* because the author's fondness for sidetracking never quite left him. However, sidetracking is the only part of the solution he really understands, so he is likely to be quite diffident about it. He feels much better about the multitude of patch mechanisms which he describes. He designed them as solutions, not problems; he wisely avoided implementing them and spoiling the illusion, so he can talk at length about how each one neatly ties up a loose end of sidetracking.

The final report usually pleases most people (more people than it should), impressing them but leaving them a little hungover. They are likely to be taken with sidetracking, especially if a theorem about it is proved, but the overall approach to the real problem lacks definition. Performance and promise run together like the colors of a sunset. The happy feeling is kindled in the reader that indefinite progress has already started. On the other hand, they usually know the author's approach won't solve everything; he avoids claiming this. So the document fails to stimulate or challenge; it merely feeds the addict's desire for reassurance that AI is not standing still, and raises his tolerance a little.

This muddle finally hurts those following in the researcher's path. Long after he has his Ph.D. or his tenure, inquiring students will be put off by the document he has left behind. He seems to have solved everything already, so the report says, yet there is no tangible evidence of it besides the report itself. No one really wants to take up the problem again, even though the original research is essentially a partial success or even a failure! If a student decides sidetracking is a good idea, and wants to study it, people will assume he is "merely implementing" an already fully designed program. (No Ph.D. for that!) He would be willing or even eager to start from a smoothly running Version II and write Version III, incorporating a new theoretical idea like Syntactic Network Data Bases, but there is no Version II. Even a Version I would help, but it isn't really working very well and its author has no desire for it to be publicized.

Of course, the student can turn his back on sidetracking, and develop an entirely new approach to Pornographic Question-Answering. But this will only antagonize people. They thought they understood sidetracking; they had convinced themselves it could be made to work. Disagreeing will only confuse them. Besides, it probably could have been made to work. If only its inventor had left it an open question!

This inflationary spiral can't go on forever. After five theses have been written, each promising with fuzzy grandeur a different solution to a problem, people will begin to doubt that the problem has any solution at all. Five theses, each building on the previous one, might have been enough to solve it completely.

The solution is obvious: insist that people report on Version I (or possibly "I½"). If a thorough report on a mere actual implementation were required, or even *allowed,* as a Ph.D. thesis, progress would appear slower, but it would be real.

Furthermore, the program should be user-engineered enough and debugged enough so that it can be run by people besides its author. What people want to know about such a program is how far they can diverge from the examples given in the thesis before it fails. Think of their awe when they discover that the hardest cases it handles weren't even mentioned! (Nowadays, the cases mentioned are, at the very best, the *only* ones the program handles.)

When a program does fail, it should tell the explorer why it failed by behavior more illuminating than, e.g., going into an infinite loop. Often a program will begin to degrade in time or accuracy before it fails. The program should print out statistics showing its opinion of how hard it had to work ("90,265 sidetracks"), so the user will not have to guess from page faults or console time. If he wishes to investigate further, a clearly written, up-to-date source program should be available for him to run interpretively, trace, etc. (More documentation should not be necessary.) In any other branch of computer science, these things are taken for granted.

My proposal is that thesis research, or any other two-year effort, should be organized as follows:

As before, a new problem, or *old problem with partial solution,* should be chosen. The part of the problem where most progress could be made (a conceptual "inner loop") should be thought about hardest. Good ideas developed here should appear in a research proposal.

The first half of the time allotted thereafter should be applied to writing Version $n+1$, where n is the version number you started with (0 for virgin problems). (Substantial rewriting of Version n should be anticipated.) The second half should be devoted to writing the report and improving Version $n+1$ with enough breadth, clean code, and new user features to make it useful to the next person that needs it.

The research report will then describe the improvements made to Version n, good ideas implemented, and total progress made in solving the original problem. Suggestions for further improvements should be included, in the future subjunctive tense.

The standard for such research should be a partial success, but AI as a field is starving for a few carefully documented failures. Anyone can think of several theses that could be improved stylistically and substantively by being rephrased as reports on failures. I can learn more by just being told why a technique won't work than by being made to read between the lines.

Benediction

This paper has focussed on three methodological and substantive issues over which we have stumbled. Anyone can think of more. I chose these because I am more guilty of them than other mistakes, which I am prone to lose my sense of humor about, such as:

1. The insistence of AI people that an action is a change of state of the world or a world model, and that thinking about actions amounts to stringing state changes together to

accomplish a big state change. This seems to me not an oversimplification, but a false start. How many of your actions can be characterized as state changes, or are even performed to effect state changes? How many of a program's actions in problem solving? (Not the actions it strings together, but the actions it *takes*, like "trying short strings first", or "assuming the block is where it's supposed to be".)

2. The notion that a semantic network is a network. In lucid moments, network hackers realize that lines drawn between nodes stand for pointers, that almost everything in an AI program is a pointer, and that any list structure could be drawn as a network, the choice of what to call node and what to call link being arbitrary. Their lucid moments are few.
3. The notion that a semantic network is semantic.
4. Any indulgence in the "procedural-declarative" controversy. Anyone who hasn't figured this "controversy" out yet should be considered to have missed his chance, and be banned from talking about it. Notice that at Carnegie-Mellon they haven't worried too much about this dispute, and haven't suffered at all. The first half of Moore and Newell (1974) has a list of much better issues to think about.
5. The idea that because *you* can see your way through a problem space, your program can: the "wishful control structure" problem. The second half of Moore and Newell (1974) is a great example.

In this paper I have criticized AI researchers very harshly. Let me express my faith that people in other fields would, on inspection, be found to suffer from equally bad faults. Most AI workers are responsible people who are aware of the pitfalls of a difficult field and produce good work in spite of them. However, to say anything good about anyone is beyond the scope of this paper.

Acknowledgment: I thank the AI Lab Playroom crowd for constructive play.

References

The numbers in square brackets following each entry indicate the chapters in which the references are cited.

[1]	Newell and Simon	[7]	Putnam
[2]	Pylyshyn	[8]	Dennett
[3]	Minsky	[9]	Haugeland
[4]	Marr	[10]	Searle
[5]	McDermott	[11]	Fodor
[6]	Dreyfus	[12]	Davidson

Abelson, R. P. (1973). "The Structure of Belief Systems." In Schank and Colby, 1973. [3]

Anderson, Alan Ross, ed. (1964). *Minds and Machines*, Englewood Cliffs, N.J.: Prentice Hall. [7] [9]

Austin, John L. (1970). *How to Do Things with Words*, New York: Oxford University Press. [Intr.]

Banerji, R., and Mesarovic, M. D., eds. (1970). *Theoretical Approaches to Non-numerical Problem Solving*, New York: Springer-Verlag. [2]

Baron, Robert J. (1970). "A Model for Cortical Memory," *Journal of Mathematical Psychology*, 7 (1970), 37–59. [9]

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*, Cambridge, England: The University Press (revised 1961). [3]

Berliner, Hans (1975). Chess as Problem Solving: The Development of a Tactics Analyzer. Unpublished Ph.D. thesis, Carnegie-Mellon University. [1]

Binford, T. O. (1971). Visual Perception by Computer. Paper delivered at the IEEE Conference on Systems and Control, Miami, Florida, December 1971. [4]

Black, Max, ed. (1965). *Philosophy in America,* Ithaca: Cornell University Press. [9]

Block, Ned, and Fodor, Jerry (1972). "What Psychological States Are Not," *Philosophical Review,* 81 (1972), 159–181. [9]

Bloomfield, L. (1933). *Language,* London: George Allen and Unwin. [11]

Bobrow, Daniel G., and Collins, A. M., eds. (1975). *Representation and Understanding,* New York: Academic Press. [5]

Bobrow, Daniel G., and Raphael, Bertram (1974). "New Programming Languages for Artificial Intelligence Research," *Computing Surveys,* 6 (1974), 153–174. [2]

Bobrow, Daniel G., and Winograd, Terry (1977). "An Overview of KRL, a Knowledge Representation Language," *Cognitive Science,* 1 (1977), 3–46. [2] [6]

Bransford, J. D., and Johnson, M. K. (1973). "Considerations of Some Problems of Comprehension." In Chase, 1973. [2]

Campbell, F. S. (1974). "The Transmission of Spatial Information through the Visual System." In Schmitt and Worden, 1974. [9]

Capitan, W. H., and Merrill, D. D., eds. (1965). *Art, Mind, and Religion,* Pittsburgh: University of Pittsburgh Press. [7]

Care, N. S. and Landesman, C. (1968). *Readings in the Theory of Action,* Bloomington, Ind.: University of Indiana Press. [8]

Carroll, Lewis (1895). "What the Tortoise Said to Achilles," *Mind,* new series, 4 (1895), 278–280 (reprinted in Copi and Gould, 1964). [8]

Castañeda, Hector Neri, ed. (1967). *Intentionality, Minds and Perception,* Detroit: Wayne State University Press. [7]

Castellan, N. John; Pisoni, David B.; and Potts, George R., eds. (1977). *Cognitive Theory II,* Hillsdale, N.J.: Lawrence Erlbaum Associates. [9]

Cathey, W. T. (1974). *Optical Information Processing and Holography,* New York: John Wiley and Sons. [9]

Cavanaugh, J. P. (1972). Holographic Processes Realizable in the Neural Realm. Unpublished Ph.D. thesis, Carnegie-Mellon University. [9]

Charniak, E. (1974). Toward a Model of Children's Story Comprehension. Unpublished Ph.D. thesis, MIT, and AI Lab Tech Report 266. [3] [5]

Chase, W., ed. (1973). *Visual Information Processing,* New York: Academic Press. [2]

Chisholm, Roderick (1967). "Intentionality." In Edwards, 1967. [8]

Chomsky, Noam (1957). *Syntactic Structures,* The Hague: Mouton. [3]

—— (1964). *Current Issues in Linguistic Theory,* The Hague: Mouton. [2]

—— (1965). *Aspects of the Theory of Syntax,* Cambridge, Mass.: MIT Press. [4]

Cohen, L. Jonathan (1950–51). "Teleological Explanation," *Proceedings of the Aristotelian Society,* 51 (1950–51), 255–292. [8]

—— (1955–56). "Can There Be Artificial Minds?" *Analysis*, 16 (1955–56), 36–41. [8]

Collins, A. W. (1969). "Unconscious Belief," *Journal of Philosophy*, 66 (1969), 667–680. [8]

Cooley, J. M. and Tukey, J. W. (1965). "An Algorithm for the Machine Computation of Complex Fourier Series," *Mathematics of Computation*, 19 (1965), 297–301. [4]

Copi, I. M. and Gould, J. A. (1964). *Readings on Logic*, New York: MacMillan. [8]

Cummins, Robert (1975). "Functional Analysis," *Journal of Philosophy*, 72 (1975), 741–765. [9]

Davidson, Donald (1970). "Mental Events." In Foster and Swanson, 1970. [9]

—— (1973a). "The Material Mind." In Suppes, et al., 1973; reprinted in this volume as Chapter 12. [9]

—— (1973b). "Radical Interpretation," *Dialectica*, 27 (1973), 313–328. [Intr.] [9]

Davies, D. J. M., and Isard, S. D. (1972). "Utterances as Programs." In Meltzer and Michie, 1972. [2]

Davis, Martin (1958). *Computability and Unsolvability*, New York: McGraw Hill. [7]

de Groot, A. (1965). *Thought and Choice in Chess*, The Hague: Mouton. [9]

Dennett, Daniel (1969). *Content and Consciousness*, London: Routledge and Kegan Paul. [8]

—— (1971). "Intentional Systems," *Journal of Philosophy*, 68 (1971), 87–106; reprinted in Dennett, 1978b, and in this volume as Chapter 8. [9]

—— (1973). "Mechanism and Responsibility." In Honderich, 1973; reprinted in Dennett, 1978b. [8]

—— (1975). "Why the Law of Effect Will Not Go Away," *Journal for the Theory of Social Behavior*, 5 (1975), 169–187; reprinted in Dennett, 1978b. [9]

—— (1976). "Conditions of Personhood." In Rorty, 1976; reprinted in Dennett, 1978b. [8]

—— (1978a). "Why You Can't Make a Computer that Feels Pain," *Synthese*, 38 (1978), 415–456; reprinted in Dennett, 1978b. [9]

—— (1978b). *Brainstorms*, Montgomery, Vt.: Bradford Books. [Intr.] [8] [9]

—— (1978c). "Two Approaches to Mental Images." In Dennett, 1978b. [9]

—— (1978d). "Toward a Cognitive Theory of Consciousness." In Savage, 1978; reprinted in Dennett, 1978b. [Intr.]

—— (1981). "Three Kinds of Intentional Psychology." In Healey, 1981. [Intr.]

Descartes, Rene (1967). *The Philosophical Works of Descartes*, Vol. 1, trans.

Elizabeth S. Haldane and G. R. T. Ross, Cambridge, England: Cambridge University Press, 1st ed., 1911. [11]

Dreyfus, Hubert L. (1972; 2nd ed., 1979). *What Computers Can't Do,* New York: Harper and Row. Excerpts from the introduction to the 2nd ed. are included in this volume as Chapter 6. [5] [6] [9]

Edwards, Paul, ed. (1967). *The Encyclopedia of Philosophy,* New York: MacMillan. [8]

Elcock, E. W.; McGregor, J. J.; and Murray, A. M. (1972). "Data Directed Control and Operating Systems," *Computer Journal,* 15 (1972), 125–129. [2]

Erickson, R. P. (1974). "Parallel 'Population' Neural Coding in Feature Extraction." In Schmitt and Worden, 1974. [9]

Ernst, G. W. and Newell, Allen (1969). *GPS: A Case Study in Generality and Problem Solving,* New York: Academic Press. [5]

Evans, T. (1968). "A Program for the Solution of Geometric-Analogy Intelligence Test Questions." In Minsky, 1968. [4]

Fahlman, Scott (1975). Thesis Progress Report: A System for Representing and Using Real-World Knowledge. Cambridge, Mass.: MIT AI Lab Memo 331. [5]

Feigenbaum, Edward A. (1977). "The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering." In *IJCAI-77,* 1014–1029. [6]

Feigenbaum, E. A., and Feldman, J., eds. (1963). *Computers and Thought.* New York: McGraw-Hill. [4]

Field, Hartry (1972). "Tarski's Theory of Truth," *Journal of Philosophy,* 69 (1972), 347–375. [11]

Firth, I. M. (1972). *Holography and Computer Generated Holograms,* London: Mills and Boon. [9]

Fodor, Jerry A. (1965). "Explanation in Psychology." In Black, 1965. [9]

—— (1974). "Special Sciences (or: The Disunity of Science as a Working Hypothesis)," *Synthese,* 28 (1974), 97–115. [9]

—— (1975). *The Language of Thought,* New York: Thomas Y. Crowell. [11]

—— (1978a). "Tom Swift and his Procedural Grandmother," *Cognition,* 6 (1978), 229–247. [11]

—— (1978b). "Propositional Attitudes," *Monist,* 61 (1978), 501–521. [11]

Fodor, Jerry, and Pylyshyn, Zenon. How Direct Is Visual Perception? Some Reflections on Gibson's "Ecological Approach." In preparation. [2]

Foster, Lawrence, and Swanson, J. W., eds. (1970). *Experience and Theory,* Amherst: University of Massachusetts Press. [9]

Frey, P., and Adesman, P. (1976). "Recall Memory for Visually Presented Chess Positions," *Memory and Cognition,* 4 (1976), 541–547. [9]

Gabor, D. (1969). "Associative Holographic Memories," *IBM Journal of Research and Development*, 13 (1969), 156–159. [9]

Geach, Peter (1957). *Mental Acts*, London: Routledge and Kegan Paul. [11]

Gibson, J. J. (1979). *An Ecological Approach to Visual Perception*, Boston: Houghton Mifflin. [2]

Gödel, Kurt (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I," *Monatshefte für Mathematik und Physik*, 38 (1931), 173–198. [Intr.]

Goldmeier, Erich (1972). *Similarity in Visually Perceived Forms*, New York: International Universities Press. [6]

Goldstein, Ira (1974). Understanding Simple Picture Programs. Unpublished Ph.D. thesis, MIT, and AI Lab Tech Report 294. [3] [5]

Goldstein, Ira, and Papert, Seymour (July 1975; revised March 1976). Artificial Intelligence, Language and the Study of Knowledge. Cambridge, Mass.: MIT AI Lab Memo 237. [6]

Goodman, Nelson (1968). *Languages of Art*, Indianapolis: Bobbs-Merrill. [9]

—— (1978). *Ways of World Making*, Indianapolis: Hackett. [11]

Grandy, Richard (1973). "Reference, Meaning and Belief," *Journal of Philosophy*, 70 (1973), 439–452. [9]

Greeno, James (1977). "Process of Understanding in Problem Solving." In Castellan, Pisoni, and Potts, 1977. [9]

Gregg, L., ed. (1974). *Knowledge and Cognition*, Potomac, Md.: Lawrence Erlbaum Associates. [2] [5]

Gregory, R. L. (1966). *Eye and Brain*, London: World University Library. [8]

Grice, H. Paul (1975). "Logic and Conversation." In Harmon and Davidson, 1975. [Intr.]

Griffiths, A. Phillips (1962–63). "On Belief," *Proceedings of the Aristotelian Society*, 63 (1962–63), 167–186; reprinted in Griffiths, 1967. [8]

—— ed. (1967). *Knowledge and Belief*, New York: Oxford. [8]

Groen, G. J. and Parkman, J. M. (1972). "A Chronometric Analysis of Simple Addition," *Psychological Review*, 79 (1972), 329–343. [2]

Gunderson, Keith, ed. (1975). *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science, 7, Minneapolis: University of Minnesota Press. [11]

Guzman, Adolfo (1968). Computer Recognition of Three-Dimensional Objects in a Visual Scene. Unpublished Ph.D. thesis, MIT, and Project MAC Tech Report 59. [6]

Hampshire, Stuart (1966). *Philosophy of Mind*, New York: Harper and Row. [8]

Harman, Gilbert (1973). *Thought*, Princeton: Princeton University Press. [9]

Harman, Gilbert, and Davidson, Donald, eds. (1975). *The Logic of Grammar,* Encino, Cal.: Dickenson. [Intr.]

Haugeland, John (1979). "Understanding Natural Language," *Journal of Philosophy,* 76 (1979), 619–632. [Intr.]

Hawkinson, L. (1975). The Representation of Concepts in OWL. Cambridge, Mass.: Project MAC Automatic Programming Group Memo 17. [5]

Healey, R. A. (1981). *Reduction, Time and Reality: Studies in the Philosophy of the Natural Sciences,* Cambridge, England: Cambridge University Press. [Intr.]

Hearst, E. (1967). "Psychology across the Chessboard," *Psychology Today,* June 1967. [9]

Heidegger, Martin (1962). *Being and Time,* New York: Harper and Row. [6]

Hempel, Carl (1962). "Rational Action," *Proceedings and Addresses of the American Philosophical Association,* 35 (1962), 5–23; reprinted in Care and Landesman, 1968. [8]

Hempel, Carl, and Oppenheim, Paul (1948). "Studies in the Logic of Explanation," *Philosophy of Science,* 15 (1948), 135–175. [9]

Herriot, D. R. (1968). "Applications of Laser Light," *Scientific American,* 218 (September 1968), 140–156. [9]

Hewitt, C. E. (1972). Description and Theoretical Analysis (Using Schemata) in PLANNER: A Language for Proving Theorems and Manipulating Models in a Robot. Cambridge, Mass.: MIT AI Lab Tech Report 258. [5]

—— (1977). "Viewing Control Structures as Patterns of Passing Messages," *Artificial Intelligence,* 8 (1977), 323–364. [2]

Hintikka, Jaakko (1962). *Knowledge and Belief,* Ithaca: Cornell University Press. [8]

Hobbes, Thomas (1651). *The Leviathan.* [Intr.]

Honderich, Ted, ed. (1973). *Essays on Freedom of Action,* London: Routledge and Kegan Paul. [8]

Hook, Sydney, ed. (1960). *Dimensions of Mind: A Symposium,* New York: New York University Press. [7]

Horn, B. K. P. (1975). "Obtaining Shape from Shading Information." In Winston, 1975. [4]

Hudson, Liam (1972). *The Cult of the Fact,* London: Cape. [9]

Hume, David (1888). *A Treatise of Human Nature,* ed. L. A. Selby-Bigge, Oxford: Oxford University Press. [11]

Husserl, Edmund (1960). *Cartesian Meditations,* The Hague: Martinus Nijhoff. [6]

IJCAI-73 (1973). Third International Joint Conference on Artificial Intelligence, *Proceedings,* Menlo Park, Cal.: SRI International, Publications Department, 333 Ravenswood Avenue, Menlo Park, Cal. 94025. [2]

IJCAI-77 (1977). Fifth International Joint Conference on Artificial

Intelligence, *Proceedings*, Available through Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Penn. 15213. [6]

James, William (1890). *Principles of Psychology*, Vol. 1, New York: Dover Publications. [11]

Jardine, N., and Sibson, R. (1971). *Mathematical Taxonomy*, New York: Wiley. [4]

Johnson, N. F. (1972). "Organization and the Concept of a Memory Code." In Melton and Martin, 1972. [2]

Julesz, B. (1975). "Experiments in the Visual Perception of Texture," *Scientific American*, 232 (April 1975), 34–43. [4]

Kabrisky, M. (1966). *A Proposed Model for Visual Information Processing in the Human Brain*, Urbana: University of Illinois Press. [9]

Kenny, K. (1963). *Trixie Belden and the Mystery of the Blinking Eye*, Racine, Wis.: Western Publishing Company. [5]

Kiefer, H. E. and Munitz, M. K., eds. (1970). *Language, Belief, and Metaphysics*, Albany, N.Y.: SUNY Press. [8]

Kosslyn, Steven, and Pomerantz, J. R. (1977). "Imagery, Propositions and the Form of Internal Representations," *Cognitive Psychology*, 9 (1977), 52–76. [9]

Krantz, D. H., et al., eds. (1974). *Contemporary Developments in Mathematical Psychology*, Vol. 2, San Francisco: W. H. Freeman. [9]

Kuhn, Thomas (1970). *The Structure of Scientific Revolutions*, 2nd ed., Chicago: University of Chicago Press. [Intr.] [3] [6] [9]

Lambert, K. (1969). *The Logical Way of Doing Things*, New Haven, Conn.: Yale University Press. [8]

Langer, S. (1962). *Philosophical Sketches*, Baltimore: The Johns Hopkins Press. [2]

Lavoisier, A. (1949). *Elements of Chemistry*, Chicago: Regnery. [3]

Leith, E. N., and Upatnieks, J. (1965). "Photography by Laser," *Scientific American*, 212 (June 1965), 24–35. [9]

Levin, M. I.; McCarthy, J.; Abrahams, P. W.; Edwards, D. J.; and Hart, T. P. (1965). *LISP 1.5 Programmer's Manual*, 2nd ed., Cambridge, Mass.: The MIT Press. [5]

Levitt, M., and Warshel, A. (1975). "A Computer Simulation of Protein Folding," *Nature*, 253 (1975), 694–698. [4]

Lewis, David (1974). "Radical Interpretation," *Synthese*, 27 (1974), 331–344. [8] [9]

Lighthill, J. (1973). "Artificial Intelligence: A General Survey." In *Artificial Intelligence: A Paper Symposium*, London: Great Britain–Science Research Council. [5]

Lucas, J. R. (1961). "Minds, Machines and Gödel," *Philosophy*, 36 (1961), 120–124; reprinted in Anderson, 1964. [Intr.]

Marr, David (1974). A Note on the Computation of Binocular Disparity in a Symbolic, Low-level Visual Processor. Cambridge, Mass.: MIT AI Lab Memo 327. [4]

—— (1976). "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society*, 275 (1976), 483–524. [4]

—— (1977a). "Analysis of Occluding Contour," *Proceedings of the Royal Society of London*, Series B, 197 (1977), 441–475. [4]

—— (1977b). "Artificial Intelligence—A Personal View," *Artificial Intelligence*, 9 (1977), 37–48; also published as MIT AI Memo 355, and included in this volume as Chapter 4. [9]

Marr, David, and Nishihara, H. K. (1977). "Representation and Recognition of Spatial Organization of Three Dimensional Shapes," *Proceedings of the Royal Society of London*, Series B, 200 (1978), 269–294. [4]

Marr, David, and Poggio, T. (1976). "Cooperative Computation of Stereo Disparity," *Science*, 194 (1976), 283–287. [4]

Martin, W. (1974). Memos on the OWL System. Cambridge, Mass.: Project MAC, MIT. [3]

Mayo, Bernard (1964). "Belief and Constraint," *Proceedings of the Aristotelian Society*, 64 (1964), 139–156; reprinted in Griffiths, 1967. [8]

McCarthy, John (1960). "Recursive Functions of Symbolic Expressions and Their Computation by Machine," *Communications of the Association for Computing Machinery*, 3 (April 1960), 184–195. [1]

—— (1979). Ascribing Mental Qualities to Machines. Stanford, Cal.: Stanford AI Lab Memo 326; reprinted in Ringle, 1979. [9] [10]

McCulloch, W. S. (1961). "What Is a Number, That a Man May Know it, and a Man, That He May Know a Number?" *General Semantics Bulletin*, Nos 26 and 27 (1961), 7–18. [1]

McDermott, Drew (1974a). Assimilation of New Information by a Natural Language-Understanding System. Cambridge, Mass.: MIT AI Lab Tech Report 291. [5]

—— (1974b). Advice on the Fast-Paced World of Electronics. Cambridge, Mass.: MIT AI Lab Working Paper 71. [5]

—— (1976). "Artificial Intelligence Meets Natural Stupidity," *SIGART Newsletter*, No. 57 (April 1976), 4–9; reprinted in this volume as Chapter 5. [6]

Melton, A. W., and Martin, E., eds. (1972). *Coding Processes in Human Memory*, New York: Winston. [2]

Meltzer, B., and Michie, D., eds. (1970). *Machine Intelligence*, Vol. 5, Edinburgh: Edinburgh University Press. [2] [3]

—— (1972). *Machine Intelligence*, Vol. 7, Edinburgh: Edinburgh University Press. [2]

Melville, Herman (1952). *Moby Dick*, New York: Modern Library College Editions. [6]

Michie, D. (1971). *On Not Seeing Things,* School of Artificial Intelligence, University of Edinburgh, EPR-22. [2]

Miller, G. A., Galanter, E., and Pribram, K. H. (1960). *Plans and the Structure of Behavior,* New York: Holt, Rinehart and Winston. [2]

Minsky, Marvin, ed. (1968). *Semantic Information Processing,* Cambridge, Mass.: MIT Press [4] [5] [6]

—— (1970). "Form and Content in Computer Science," *Journal of the Association for Computing Machinery,* 17 (January 1970), 197–215. [3]

—— (1974). A Framework for Representing Knowledge. Cambridge, Mass.: MIT AI Lab Memo 306; excerpts reprinted in Winston, 1975; other excerpts reprinted in: *TINLAP-75;* and still others included as Chapter 3 of this volume. [4] [6]

Minsky, Marvin, and Papert, Seymour (1970). Draft of a proposal to ARPA for research on artificial intelligence at MIT, 1970–71. [6]

—— (1972). Progress Report on Artificial Intelligence. Cambridge, Mass.: MIT AI Lab Memo 252. [3] [9]

—— (1973). *Artificial Intelligence,* Condon Lectures, Oregon State System of Higher Education, Eugene, Oregon. [6]

Moore, J., and Newell, Allen (1974). "How Can Merlin Understand?" In Gregg, 1974. [2] [5]

Moses, J. (1974). "MACSYMA—The Fifth Year," *SIGSAM Bulletin,* 8 (1974), 105–110. [4]

Nevins, A. (1974). A Relaxation Approach to Splitting in an Automatic Theorem Prover. Cambridge Mass.: MIT AI Lab Memo 302. [5]

Newell, Allen (1962). "Some Problems of Basic Organization in Problem-Solving Programs." In Yovitts, Jacobi, and Goldstein, 1962. [2] [5]

—— (1970). "Remarks on the Relationship between Artificial Intelligence and Cognitive Psychology." In Banerji and Mesarovic, 1970. [2]

—— (1972). "A Theoretical Exploration of Mechanisms for Coding the Stimulus." In Melton and Martin, 1972 [2]

—— (1973a). "Artificial Intelligence and the Concept of Mind." In Schank and Colby, 1973. [3]

—— (1973b). "You Can't Play 20 Questions with Nature and Win." In Chase, 1973. [2]

—— (1973c). "Production Systems. Models of Control Structures." In Chase, 1973. [2]

—— (1980). Physical Symbol Systems. *Cognitive Science,* (1980), 135–183. [10]

Newell, Allen, and Simon, Herbert (1972). *Human Problem Solving,* Englewood Cliffs, N.J.: Prentice-Hall. [2] [3]

Nilsson, Nils J. (1971). *Problem Solving Methods in Artificial Intelligence,* New York: McGraw Hill. [1]

Norman, Donald (1973). "Memory, Knowledge and the Answering of Questions." In Solso, 1973. [3]

Norman, D. A., and Rumelhart, D. E. (1974). *Explorations in Cognition*, San Francisco: W. H. Freeman. [4]

Papert, Seymour (1972). "Teaching Children to be Mathematicians vs. Teaching about Mathematics," *International Journal of Mathematical Education for Science and Technology*, 3 (1972), 249–262. [3]

Papert, Seymour, and Minsky, Marvin (1973). Proposal to ARPA for Research on Intelligent Automata and Micro-Automation. Cambridge, Mass.: MIT AI Lab Memo 299. [6]

Pavio, Allen (1975). "Imagery and Synchronic Thinking," *Canadian Psychological Review*, 16 (1975), 147–163. [9]

Poggio, T., and Reichardt, W. (1976). "Visual Control of the Orientation Behavior of the Fly: Towards the Underlying Neural Interactions," *Quarterly Reviews of Biophysics*, 9 (1976), 377–438. [4]

Pollen, D. A., and Taylor, J. H. (1974). "The Striate Cortex and the Spatial Analysis of Visual Space." In Schmitt and Worden, 1974. [9]

Pribram, Karl H. (1971). *Languages of the Brain*, Englewood Cliffs, N.J.: Prentice-Hall. [9]

—— (1974). "How Is It That Sensing So Much We Can Do So Little?" In Schmitt and Worden, 1974. [9]

Pribram, Karl H.; Nuwer, M.; and Baron, Robert J. (1974). "The Holographic Hypothesis of Memory Structure in Brain Function and Perception." In Krantz, et al., 1974. [9]

Price, H. H. (1954). "Belief and Will," *Proceedings of the Aristotelian Society*, Supplementary Vol. 28 (1954), 1–26; reprinted in Hampshire, 1966. [8]

Putnam, Hilary (1960). "Minds and Machines." In Hook, 1960; reprinted in Anderson, 1964, and in Putnam, 1975b. [7] [9] [11]

—— (1965). "Psychological Predicates." In Capitan and Merrill, 1965. [7]

—— (1967). "The Mental Life of Some Machines." In Castañeda, 1967; reprinted in Putnam, 1975b. [7]

—— (1973). "Reductionism and the Nature of Psychology," *Cognition*, 2 (1973), 131–146; abridged version included in this volume as Chapter 7. [9]

—— (1975a). "The Meaning of 'Meaning.'" In Gunderson, 1975; reprinted in Putnam, 1975b. [11]

—— (1975b). *Mind, Language and Reality–Philosophical Papers*, Vol. 2, Cambridge, England: Cambridge University Press. [7] [9] [11]

Pylyshyn, Zenon (1973). "What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery," *Psychological Bulletin*, 80 (1973), 1–24. [9]

—— (1978a). "Imagery and Artificial Intelligence." In Savage, 1978. [9]

—— (1978b). "Computational Models and Empirical Constraints," *Behavioral and Brain Sciences*, 1 (1978), 93–99. [2]

—— (1980). "Computation and Cognition: Issues in the Foundations of Cognitive Science," *Behavioral and Brain Sciences*, 3 (1980), 111–132. [2]

Quillian, M. R. (1968). "Semantic Memory." In Minsky, 1968. [4] [5]

—— (1969). "The Teachable Language Comprehender," *Communications of the Association for Computing Machinery*, 12 (1969), 459–476. [5]

Quine, W. V. O. (1960). *Word and Object*, Cambridge, Mass.: MIT Press. [Intr.] [8] [9]

Raphael, B. (1968). "SIR: Semantic Information Retrieval." In Minsky, 1968. [4]

Ringle, Martin, ed. (1979). *Philosophical Perspectives in Artificial Intelligence*, Atlantic Highlands, N.J.: Humanities Press. [2] [9] [10]

Robinson, J. A. (1965). "A Machine-Oriented Logic Based on the Resolution Principle," *Journal of the Association for Computing Machinery*, 12 (1965), 23–41. [5]

Rorty, Amelie O., ed. (1976). *The Identities of Persons*, Berkeley: University of California Press. [8]

Rosch, Eleanor (1976). "Classifications d'objets du monde réel: origines et représentations dans la cognition," *Bulletin de Psychologie* (1976), 242–250. [4]

—— (1977). "Human Categorization." In Warren, 1977. [4] [6]

Sacerdoti, E. D. (1973). "Planning in a Hierarchy of Abstraction Spaces." In *IJCAI-73*, 412–422; also published in *Artificial Intelligence*, 5 (1974), [2]

Sandewall, E. (1970). "Representing Natural Language Information in Predicate Calculus." In Meltzer and Michie, 1970. [3]

Savage, C. Wade, ed. (1978). *Perception and Cognition: Issues in the Foundations of Psychology*, Minnesota Studies in the Philosophy of Science, Vol. 9, Minneapolis: University of Minnesota Press. [Intr.] [9]

Schank, Roger C. (1972). "Conceptual Dependency: A Theory of Natural Language Understanding," *Cognitive Psychology*, 3 (1972), 552–631. [6]

—— (1973a). The Fourteen Primitive Actions and Their Inferences. Stanford, Cal.: Stanford AI Lab Memo 183. [3]

—— (1973b). "Identification of Conceptualizations Underlying Natural Language." In Schank and Colby, 1973. [4]

—— (1975a). "The Primitive Acts of Conceptual Dependency." In *TINLAP-75*. [6]

—— (1975b). "Using Knowledge to Understand." In *TINLAP-75*. [6]

—— (1975c). *Conceptual Information Processing*, Amsterdam: North-Holland. [4]

—— (1979). "Natural Language, Philosophy and Artificial Intelligence." In Ringle, 1979. [2]

Schank, Roger C., et al. (1977). "Panel on Natural Language Processing." In *IJCAI-77*, pp1007–1013. [6]

Schank, Roger C., and Ableson, Robert P. (1977). *Scripts, Plans, Goals and Understanding,* Hillsdale, N.J.: Laurence Erlbaum Associates. [6] [10]

Schank, Roger C., and Colby, Kenneth, eds. (1973). *Computer Models of Thought and Language,* San Francisco: W. H. Freeman. [3] [4] [6]

Schmitt, F. O., and Worden, F. G., eds. (1974). *The Neurosciences: Third Study Program,* Cambridge, Mass.: MIT Press. [9]

Searle, John R. (1969). *Speech Acts: An Essay in the Philosophy of Language,* Cambridge, England: Cambridge University Press. [Intr.]

—— (1979). "What Is an Intentional State?" *Mind,* 88 (1979), 72–94. [10]

Sellars, Wilfrid (1963). *Science, Perception and Reality,* London: Routledge and Kegan Paul. [2] [9]

Shepard, Roger (1975). "Form, Formation, and Transformation of Internal Representations." In Solso, 1975. [4]

Shepard, Roger, and Metzler, J. (1971). "Mental Rotation of Three-Dimensional Objects," *Science,* 171 (1971), 701–703. [9]

Shoemaker, Sydney (1975). "Functionalism and Qualia," *Philosophical Studies,* 27 (1975), 291–315. [9]

Simmons, R. F. (1973). "Semantic Networks: Their Computation and Use for Understanding English Sentences." In Schank and Colby, 1973. [3]

Simon, Herbert A. (1969). *The Sciences of the Artificial,* Compton Lectures, Cambridge, Mass.: MIT Press. [2] [9]

—— (1977). "Artificial Intelligence Systems That Understand." In *IJCAI-77,* 1059–1073. [6]

Slagle, J. R. (1963). "A Heuristic Program That Solves Symbolic Integration Problems in Freshman Calculus." In Feigenbaum and Feldman, 1963. [4]

Solso, R. L., ed. (1973). *Contemporary Issues in Cognitive Psychology: The Loyola Symposium,* Washington, D.C.: V. H. Winston and Sons. [3]

—— , ed. (1975). *Information Processing and Cognition: The Loyola Symposium,* Hillsdale, N.J.: Lawrence Erlbaum Associates. [4]

Stansfield, J. L. (1975). Programming a Dialogue Teaching Situation. Unpublished Ph.D. thesis, University of Edinburgh. [5]

Sternberg, S. (1967). "Two Operations in Character Recognition," *Perception and Psychophysics,* 2 (1967), 43–53. [2]

Sunguroff, A. (1975). Unpublished paper on the OWL system. [5]

Suppes, Pat, et al., eds. (1973). *Logic, Methodology and Philosophy of Science IV,* Amsterdam: North-Holland. [9]

Sussman, Gerald J. (1973). *A Computational Model of Skill Acquisition.* MIT Ph.D. thesis and AI Lab Tech Report 297; published by American Elsevier, New York, 1975. [3] [5]

Sussman, Gerald J., and McDermott, Drew V. (1972). "From PLANNER to CONNIVER—A Genetic Approach (Or: Why Conniving Is Better Than Planning)," *AFIPS Conference Proceedings,* 41 (1972), 1171–1180. [5]

Sussman, G. J., and Stallman, R. M. (1975). "Heuristic Techniques in Computer-Aided Circuit Analysis," *IEEE Transactions on Circuits and Systems, CAS-22* (1975), 857-865. [4]

Taylor, Charles (1964). *The Explanation of Behavior,* London: Routledge and Kegan Paul. [8]

TINLAP-75 (1975). *Theoretical Issues in Natural Language Processing,* Cambridge, Mass.: June 10-13, 1975. [6]

Turing, A. M. (1937). "On Computable Numbers, with an Application to the *Entscheidungsproblem*," *Proceedings of the London Mathematical Society,* 42 (1937), 230-265. [Intr.] [2]

—— (1950). "Computing Machinery and Intelligence," *Mind,* 59 (October 1950), 433-460; reprinted in Anderson, 1964. [1] [10]

Ullman, S. (1976). "On Visual Detection of Light Sources," *Biological Cybernetics,* 21 (1976), 205-212. [4]

van Heerden, P. J. (1963). "A New Method of Storing and Retrieving Information," *Applied Optics,* 2 (1963), 387-392. [9]

Vickers, John (1969). "Judgement and Belief." In Lambert, 1969. [8]

Waltz, David (1972). *Generating Semantic Descriptions from Drawings of Scenes with Shadows.* MIT Ph.D. thesis; published in Winston, 1975. [4] [6] [9]

Warren, N., ed. (1977). *Advances in Cross-Cultural Psychology,* Vol. 1, London: Academic Press. [6]

Warrington, E. K. (1975). "The Selective Impairment of Semantic Memory," *Quarterly Journal of Experimental Psychology,* 27 (1975), 635-657. [4]

Weizenbaum, Joseph (1965). "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the Association for Computing Machinery,* 9 (1965), 36-45. [4] [10]

—— (1976). *Computer Power and Human Reason,* San Francisco: W. H. Freeman. [4] [5] [10]

Wertheimer, M. (1959). *Productive Thinking,* New York: Harper and Row. [3]

Williams, B. A. O. (1970). "Deciding to Believe." In Kiefer and Munitz, 1970. [8]

Wilson, N. L. (1959). "Substances without Substrata," *Review of Metaphysics,* 12 (1959), 521-539. [9]

Winograd, S. (1976). "Computing the Discrete Fourier Transform," *Proceedings of the National Academy of Science,* 73 (1976), 1005-1006. [4]

Winograd, Terry (1971). Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. Cambridge, Mass.: MIT AI Lab Tech Report 84. [5] [11]

—— (1972). "Understanding Natural Language," *Cognitive Psychology,* 1

(1972), 1–191; also published by Academic Press, New York, 1972. [2] [6] [9] [10]

—— (1973). "A Procedural Model of Language Understanding." In Schank and Colby, 1973. [6]

—— (1974). Five Lectures on Artificial Intelligence. Stanford, Cal.: Stanford AI Lab Memo 246. [3] [6]

—— (1975). "Frame Representations and the Declarative/Procedural Controversy." In Bobrow and Collins, 1975. [5]

—— (1976a). "Artificial Intelligence and Language Comprehension." In *Artificial Intelligence and Language Comprehension,* Washington, D.C.: National Institute of Education. [6]

—— (1976b). "Towards a Procedural Understanding of Semantics," *Revue Internationale de Philosophie,* Nos. 117–118 (1976), 260–303, Foundation Universitaire de Belgique. [6]

Winston, Patrick H., ed. (1975). *The Psychology of Computer Vision,* New York: McGraw Hill. [4] [6]

Winston, Patrick H., and the staff of the MIT AI Laboratory (May, 1976). Proposal to ARPA, Cambridge, Mass.: MIT AI Lab Memo 366. [6]

Wittgenstein, Ludwig (1953). *Philosophical Investigations,* Oxford: Basil Blackwell. [6]

Woods, William A. (1975). "What's in a Link: Foundations for Semantic Networks." In Bobrow and Collins, 1975. [5]

Yevick, Miriam L. (1975). "Holographic or Fourier Logic," *Pattern Recognition,* 7 (1975), 197–213. [9]

Yovitts, M.; Jacobi, G. T.; and Goldstein, G. D. (1962). *Self-Organizing Systems,* New York: Spartan. [2] [5]

Zucker, S. W. (1976). Relaxation Labeling and the Reduction of Local Ambiguities. College Park, Md.: University of Maryland Computer Science Tech Report 451. [4]