TECHNICAL REPORT NO. 176

WHY ARTIFICIAL INTELLIGENCE IS NECESSARILY AD HOC:
One's Thinking/Approach/Model/Solution
Rides on One's Metaphors

by

Marek W. Lugowski

August 1985

COMPUTER SCIENCE DEPARTMENT

INDIANA UNIVERSITY

Bloomington, Indiana 47405-4101

WHY ARTIFICIAL INTELLIGENCE IS NECESSARILY AD HOC:

One's Thinking/Approach/Model/Solution Rides on One's Metaphors

Marek W. Lugowski

Indiana University Computer Science Department
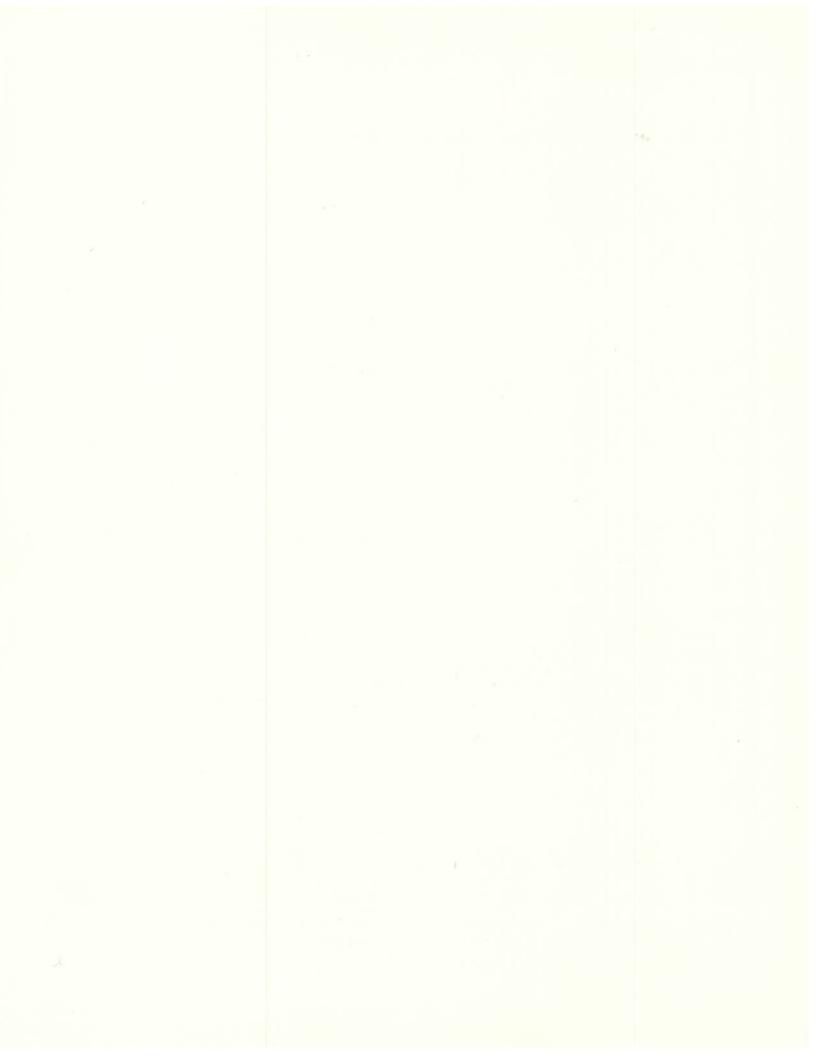Bloomington, Indiana 47405

WHY ARTIFICIAL INTELLIGENCE IS NECESSARILY AD HOC:

One's Thinking/Approach/Model/Solution Rides on One's Metaphors

Marek W. Lugowski

Indiana University Computer Science Department
Bloomington, Indiana 47405

TECHNICAL REPORT NO. 176

WHY ARTIFICIAL INTELLIGENCE IS NECESSARILY AD HOC:
One's Thinking/Approach/Model/Solution
Rides on One's Metaphors

by

Marek W. Lugowski

August 1985

COMPUTER SCIENCE DEPARTMENT

INDIANA UNIVERSITY

Bloomington, Indiana 47405-4101

WHY ARTIFICIAL INTELLIGENCE IS NECESSARILY AD HOC:

One's Thinking/Approach/Model/Solution Rides on One's Metaphors

Marek W. Lugowski

Indiana University Computer Science Department
Bloomington, Indiana 47405
Technical Report No. 176
August 1985

ABSTRACT

The *ad hoc* label is a derogatory one in the artificial intelligence (AI) community. It is customarily bestowed on poorly thought-out work that lacks wider applicability beyond its immediate concern. It is applied to work deemed to lack *systematicity*, a quality nowadays equated with structured methodologies a la David Marr.

The issue of the ad hoc versus the systematic is reexamined. It is claimed that even the most highly regarded (systematic) AI research is often a product of inferior theory-formation. The weakest point of to-date AI research is the thinking up of metaphors in the course of problem conception.

A claim is made that metaphorizing is ad hoc in the sense that there is no way to systematize the process of conjuring, selecting, and using metaphors. In fact, it is claimed that AI theory-formation ought to transcend whatever systematicity may be found in everyday metaphor use, and perhaps even include metaphor mixing.

The alternative, presently in vogue within AI, that of running a single metaphor into the ground when developing a conceptualization or discussing issues, is perceived to cause skewed perception, inadequate problem conception, and undue bias in implementation.

Ad hoccery inherent in metaphorizing suggests that structured methodologies stifle AI and subvert much of AI's theory-formation. Namely, (1) it is impossible to cleanly separate the theory being thought from the thinker's conceptualizing tools (including implementation metaphors), and (2) AI research involves unconstrained introspection to a degree not found anywhere else in engineering or science. Therefore, to force-fit AI thinking into a mold roughly equivalent to a set-theoretical partitioning into unrelated levels of abstraction begs the complex question of one's awareness of one's theory-formative thinking.

Good metaphor use is also cited as needed for effective research reporting. Four cases of well-received projects are presented in order to support the thesis that AI metaphorizing needs more attention if AI goals are to be achieved.

August 1, 1985

I. Prolegomena to the Rehabilitation of the Ad Hoc

I.1 The ad hoc versus the systematic

A recurring theme in college AI instruction is that of warning students about ad hoc work in artificial intelligence (AI): [WINSTON 83a], [DAVIS 83]. *Ad hoc* is a funny term. When used in contexts such as "The Senate's Ad Hoc Committee on ...", it evokes a sight of concerned lawmakers, dropping their usual pork-barrel politics to supplely take up a pressing matter. Yes, ad hoc can evoke images of Yankee ingenuity, resolute and apt action, and a certain air of fresh, inspiring gestalt perceptions. Surely that is not what is meant by ad hoc in the AI lectures I recall, judging from context. There, ad hoc is an epithet heaved upon work perceived to be callously patched together in an unjustified (unjustifiable?) fashion, or, to put it in MITese, a "crufty hack".

The antonym of ad hoc in AI phenomenology is "the systematic". David Marr's posthumous [MARR 82] is regarded highly not only for its technical contributions, but also for its systematicity. Marr's insistence on layering the process of AI research in clearly distinct strata and his insistence on the systematic spelling-out of competences a project is to achieve have made him something of a modern (and vogueish) Thomas Aquinas at the MIT AI Lab.

This paper has little to do with Marr's contributions to research in vision. But his fabulously good press and good standing as the Paragon of AI-ship in Form and Substance is a good point of departure for a closer look at the field's self-ascribed wretchedness and true sins: items, I claim, not quite so synonymous as the Inquisition would have it.

I.2 An ad hoc view of systematicity (read 'ad hoc' both ways)

Speaking of fabulously good press, the AI loci of activity surrounding the study of analogy/metaphor, linguistic and not, could not be held in higher regard than they are. The plurality opinion is that analogizing and various metaphorical phenomena pervade, if not account for, the bulk of the cognitive (to some, subcognitive) computation. In this age of heightened international tension and ideological struggle, one is comforted to see any view be given fabulously good press, let alone a view so excitingly romantic.

In light of what I have said so far, I would like to consider the implications of this triad of ideas:

> 1. Ad hoc AI is bad AI.
> 2. Metaphor underlies (is) everything (sub)cognitive.
> 3. AI is chiefly insight-driven.

The first two ideas have already been mentioned. The third is floated just as often as the other two, judging from the public remarks made by such AI standouts as Roger Schank and Marvin Minsky, for example, in [SCIENCE85] and [MINSKY]. I have it on good rumor that most AI researchers subscribe to Number 3, though they may choose to neither say so in their submissions to the various neuro/linguo/psychojournals nor flaunt it in front of their dissertation or tenure-granting committees. After all, to outsiders (and, apparently, some insiders) the stuff of AI is science (or engineering), and even the "cats clanking with coin" at D.O.D. (apologies to Joni Mitchell) *know* that science is object/subject, and that good science always reveals the same immutable reality. (However, [CODE 85] and [KELLER 85] beg to differ.)

I am now proof-theoretically in a position to reveal the promised implications:

> 4. Metaphor use in thinking about AI determines AI. (by 2, 3)
>
> 5. Metaphor use determines ad hoc and systematic AI. (by 4)
>
> 6. Metaphor use is ad hoc. (by Dreyfusian Intuition ([DREYFUS 79],
>   [SCIENCE85]) and into a strange-loop ([HOFSTADTER 79])!)

There! Tongue-in-cheek, perhaps, and maybe lacking in dignity, but let those with logs in their eyes cast the first tomato of refutation [BUSH] at No. 6. (Mixing metaphors does not cause blindness or insanity.) I daresay that even Marr never metaphorized systematically, nor did he leave an effective procedure for doing so (not even a systematic description of such competence) amongst his two-and-a-half-dee sketches. I wish to claim that both the metaphor creation/selection process *and* the actual metaphor use —by everybody —are ad hoc.

The inner systematicity of a metaphor is a very weak constraint on the thinker. Contrary to [GENTNER 83] and [LAKOFF & JOHNSON 80], the constraints of idiomatic metaphor use —at least in English and Polish! —allow an infinity of possibilities. The less-frequented regions of the metaphor-use space are interpreted by us as manifesting creativity, in contrast to those which are employed as everyday liguistic currency. This leads me to suspect that creativity is a matter of societal perception, not so much an individual characteristic. It is resourcefulness that deserves our reverence and our AI efforts instead. Therefore, to seek guidance from the inner systematicity of a common metaphor, be it in theory-formation or in "structural analogies", is no more constraining than relying on the semantics of a first-order logic to shape declarative entities, e.g., semantic networks. [GENTNER 83] and [WINSTON 80-82] sing praises of the former (metaphor systematicity), but [HAYES 77] and [HAYES 83] eloquently scorn the latter (logic vs. declarative knowledge constraints). As disinterested in logic as I am, I cannot resist making a snide remark on this inconsistency under a logical

equivalence — and all this flying-in-the-face within a mutual admiration club, at that! I itch to see this sort of reflection handled by the mighty tower of layered interpreters and meta-meta logic that is now taking the worlds of "no-nonsense" AI and programming languages research by storm ([BATALI 83], [MCCARTHY 80], [SMITH 82], [WEYROUCH 80]).

The systematic exploration of an idea within the confines of a single metaphor's systematicity has a way of painting the AI theorizer into a corner. I will say more on that later. As for creating and selecting metaphors from one's repertoire of experiences and images, who can provide but the vaguest of guide-lines? About the best way to give oneself healthy AI theory-formation skills is to obtain a great variety of experiences...non-traumatizing ones. (One should only go so far for the sake of professional self-improvement!)

By the way, does not my inference chain (1. through 6.) remind you of AI research as such?: a quaint mix of the formal with the whimsical, seasoned with the whimsical insistence on formalizing (in logic, please!) everything whimsical in sight. Be that as it may, and all pomp and circumscription [MCCARTHY 80] aside, Number 6 is true.

If I were an unfriendly philosopher from Berkeley ([DREYFUS 79], [SEARLE 82]), I would proceed to derive that ad hoccery is good AI, thus a contradiction, and poof! —purport to expose the entire shadowy business as inconsistent, circu-lar, and thus unformalizable. Obligatorily, I would then exhort it to crawl back under the rock (of mathematics), where it came from.

However, I'm not from Berkeley. Furthermore, my intentions are not to taunt AI, but merely to continue within it, mending some of its traditions, and improving its way of doing business. For it is my hunch that:

> 7. The most systematic thinking that AI can muster must be initiated and nurtured by certain consummately ad hoc thinking skills, namely, ample metaphorizing.

I.3 Theory formation deficiencies of established research

To back up Number 7, my thesis, I offer four case studies of metaphor use and abuse (a.k.a. theory formation) in known AI research. I chose the cases with specific, but different, metaphorizing faults in mind, and without particular correlation to perceived ad hoccery. This means that the presented work is in good standing with sizable minorities. That's about the best one can do given the present state of the (in)discipline.

The faults, usually related to insufficient ad hoccery at the time of initial problem conception, that is, metaphorizing, are implicated as the causes of, to put it tactfully, imperfect research or, arguably, merely imperfect research reporting. Either way, one rarely has the luxury of discerning one from the other, and often the distinction is academic: Routinely, entirely sanguine thinkers forgo doubting their certitudes, once cast in print.

The research discussed herein is [WINSTON 79], [LENAT 82] and [LENAT 83], [TURKLE 84], and [HOFSTADTER 84], in that order. Turkle may fight her extradition from the sphere of sociology and psychology, but her questions (and mistakes) are echoed throughout AI daily.

Before I begin, I should like to note that it is not my intention to seek banishment from the AI community (or worse, to be disrespectful of my teachers and elders) by panning esteemed work. Perhaps disdain, or, more likely, a lack of notice and plain indifference may be the greeting for this essay, but such is par for the course of scholarship.

My critical look at the metaphors of respected researchers would not have been possible had these people left alone the difficult issues that got them into trouble. It is much easier to come later and notice how a person's metaphorizing influences other people's, and how it channels the originator's own new ideas. With this thought in mind, I am driven to stress the need for paying fastidious attention to one's theory-formative hygiene: minding one's metaphors, greedily acquiring new ones, and occasionally mixing them to great profit.

II. The Case of Robbie Who Is Like a Fox: Winston's Transfer-Frames

In this case study I call into question Patrick H. Winston's work on "learning through simile", and by implication, the broader issue of using his ARCH-introduced pragmatics for tasks beyond the conjuring of coarse object descriptions as "a naive algebra" (apologies to Pat Hayes) on very simple and small graphs. I perceive the faults in his theory-formative thinking as (1) insufficient, even stunted, metaphorizing, and relatedly (2) much too much reliance on his prior approach. While continuity is praiseworthy in ethics and politics, chronic attachment to one's doctoral dissertation in turbulent, insight-driven AI can stultify and lead one onto garden paths. New conceptualizations remain inaccessible to thinkers such as Winston because their *new* ideas are shaped by unstated limitations inherited from their *old* work, especially their old implementations. Three such limitations in Winston's learning work are rigid top-down sequentiality, rigid parsing-as-inference conceptualizations that strongly constrain the nature of proposed solutions, and rigid treatment of symbols and other fluid, emergent phenomena of message-passing —as static, atomic Lisp entities.

Winston is perceived by most as quite systematic in his work, which is as expected given his explicit endorsement and adaptation of Marr's approach. Yet, I smirk every time I read a certain passage in [WINSTON 79] which deals with learning from sentences such as *Robbie is like a fox*. More importantly, it elucidates Winston's general approach. As for *Robbie is like a fox*, we are to look at it, which will make us look at *Robbie*, then at a *fox*, and then we are to understand that *Robbie* is as clever as a *fox*.

In Winston's transfer-frame world, everything fits neatly into a nice hierarchy that forms a tree-network. Furthermore, everything is a bundle of attribute/value pairs, where each one is a nice, simple, short word. When we are presented with a "simile", we retrieve the two bundles. Then, a certain number

of candidate transfer-frames are created by screening the "fox" bundle against its immediate relatives, i.e., siblings in the world-tree. These prospective transfer-frames are to act as sources of potential transfers of attribute values from the "fox" part to the "Robbie" part. This process constitutes "learning through simile". Before anything is learned, however, "the better" of the transfer-frames are chosen after doing a background check on "Robbie" and his close world-family, and after remembering any previous learning. Finally, one lucky transfer-frame remains salient and is used to bring about the learning.

The sentence I find amusing attempts to deal with the issue of establishing context for sentences such as *Robbie is like a fox*. Winston reduces this issue to the matter of consulting the "immediate relatives" of a fox in order to permit us to glean the cleverness attributed to Robbie. This is supposed to work in the following way: If a *fox* has an attribute that is thought to be particularly salient (whether this salience is momentary or permanent is not addressed by Winston), and moreover, if that attribute is not similarly salient for *fox*'s neighbors in the world-space (on the assumption that such a space can or should be constructed!), then one is free to conclude that the sentence *Robbie is like a fox* must be drawing our attention to the fact that Robbie ought to be endowed with the same attribute. The process of making this conclusion and endowing Robbie's matching attribute with a suitable value is what Winston wishes to be known as "learning through simile". Consequently, the sentence that I find so amusing reads as follows: "For a fox, other small common forest mammals would be used."

This makes perfect sense to Winston, to whom a fox is "A-KIND-OF SMALL-COMMON-FOREST-MAMMAL", and furthermore, who bestows upon it a salient attribute (according to a byzantine salience-divining scheme) of "CLEVERNESS" with a corresponding value of "VERY-HIGH", which the other SMALL-COMMON-

FOREST-MAMMAL thingees have not. Thus, presto, he seizes on the fact that this "CLEVERNESS" is what is being got across, and he rests his case, having "learned" Robbie's cleverness.

Why are "other small common forest mammals" the immediate relatives of a fox, however? Are foxes always small common forest mammals? Would this be so in *Robbie is a fox*? Wouldn't then Robbie just as likely be thought of as a sassy, good-looking woman — alternatively, the speaker to be a male sexist pig (or both, or neither, depending on one's politics)? Since when did words, or, strictly speaking, the typographical manifestations of meaning become well-bundled, forever-determined entities with equally static immediate relatives, regardless of usage, whatever the context, pun or no pun? An objection could be made that *Robbie is like a fox* is a far cry from *Robbie is a fox*, the latter of which does not even seem to be a simile. I intend to answer this objection below when I discuss the nature of similes, but for now, consider the sentence *Sally is a block of ice* and its companion *Sally is like a block of ice*. Clearly, it doesn't matter much that the term "like" is absent from one of those sentences. Both say roughly the same thing. Continuing onto to a higher level of abstraction, let us muse over: *The-meaning-of-Sally-sentences is like The-meaning-of-Robbie-sentences* — yes or no? After weighing our answers, let us take on *Robbie is like a cuckoo*.

Are we going to look at thrushes, redbirds, and assorted woodpeckers and somehow fathom that Robbie is "crazy"? Imagine there, under CUCKOO (is) A-KIND-OF BIRD, an attribute to the effect: FIGURATIVE-PEJORATIVE-SIGNIFICANCE-AS-CRAZY-IN-THE-HEAD (BUT-ONLY-IN-ENGLISH-AND-LATELY-NOT-USED-OFTEN) with a corresponding value VERY-HIGH. Who would create these all-describing attributes? Who would choose one over another? Who would adjust their relative importance during "learning"? Who would keep the world in a nice tree? Is it possible to do that? What happens to the state of the world

as soon as "learning" is achieved? Can change adequately propagate from one slot to the other slots that depend on its value, in a rigid world? Is it possible to maintain consistency, dependency, locality, subjectivity, and saliency —as well as represent conflicting conceptions, attitudes, certitudes, the passing of time, and fantasizing —in this static world-tree? Is this scheme good enough to describe even a small, changing, self-contained world and handle "learning through simile"? How does it generalize to open worlds? Is there really any learning taking place? Are there any similes here?

In case you're wondering about that last one, notice that the system Winston describes does not accept similes as input. A simile is a literary device whose lexical manifestation in English typically contains the term "like", and only if the prose does not originate from San Fernando Valley, California, for then the "like" is, like, redundant. Noticing the "like" in a simile is like a single ant on the anthill of defining or understanding the process of simile —to use a simile like a Rose Bowl. Winston is not even working on a piece of natural language syntax, let alone investigating a deep semantic idea of what simile is and how it works. What Winston's system accepts constitutes ordered pairs of items which are immediately pigeonholed as "a source" and "a destination".

If Winston says he is dealing with similes, then logical equivalence dictates that Skinner's pigeons [SKINNER] must be said to be as well, as neither Winston nor the pigeons appear to show any inkling of defining or understanding the process of simile.

In the case of Winston, his program crudely connects with functional arcs the nodes representing the components of the purported simile, as if they were blocks comprising one of his famous arches. In fact, Winston's ARCH and TRANSFER-FRAMES implementations and descriptions overlap greatly. One shouldn't praise him for simulating similes with this simple "putpropy" associa-

tionism; he is merely thinking by analogy with his former, simpler, and —this is crucial —vastly different topic.

Apparently to Winston a simile is just A-KIND-OF ARCH, with links and nodes suitably painted over. The thrust of my taking Winston to task here is this innocent "if similes were arches" thinking of his, which nota bene, lends a marvelously new point of departure for analyzing that famous Lewis Carroll morsel about wishes, horses, and beggars...

In the case of Skinner's pigeons, their conditioning involves making associations between colored squares in a learned and useful way (that is, yielding grub). But really, isn't Skinner hacking the same association making (he calls it "behavior modification", "conditioning") that Winston calls "learning by simile"? I think that Marr would say that the computational theory and algorithms are identical in both cases: It is either Skinner who has implemented Winston's "similes" in pigeons, or it is Winston who is hacking pigeonhood into Lisp. Whatever, the birds peck wisely at the provided colored squares, making associations that are nonliterary only because there ain't no "fox" or "Robbie" written on them squares (apologies to John Searle), if we buy Winston's self-styling as simile-hacker. (As a matter of fact, some pecks were directed at buttons marked "Thank you" and "What color?". Can we accuse the pigeons of "verbal written communication"???) The bottom line is that pigeons learn to peck "through simile" in very complex ways, in fact, through operating machinery and using symbols, committing precisely those peck-acts (apologies to Roger Schank and John Searle) that reward them. Incidentally, the Harvard pigeon is somewhat cheaper to get and maintain than Winston's DecSystem-20 with Maclisp, but then MIT is funded by the Office of Naval Research, so who knows what a General Dynamics-contracted pigeon goes for these days... But surely even a very expensive D.O.D. pigeon doesn't *really* learn through simile. Or does it?

And what about my *Robbie is like a cuckoo*? I don't mean to suggest that Winston should've written up a project to simultaneously handle English, French, and Moldavian, but as a convenient example of the sort that can be found in any language, consider that in Polish a cuckoo is thought of as an opportunistic dude and a lousy parent notoriously dumping its parental obligations into some dumb unsuspecting birdie's nest (not a cuckoo's!). Also, even little Poles know that juvenile cuckoos are selfish, egotistic hooligans that often crowd their foster siblings out of the nest in a grim pursuit of room and board. In Poland —as much as in America —*Robbie is like a cuckoo* could well elicit a swift fist from some disapproving Robbie, especially if the simile happened to be on target. On the other hand, to liken a Pole who is running around a lot to a cuckoo could evoke a tired smile, as intended, for the kind analogy to the to-and-fro wooden clock figurine can be conveyed idiomatically. But try putting that one over on a Polish journalist, or a parish priest: Neither the journalist, who has enough of a job in Poland trying to appear a priori credible over and above the usual trivia that journalists the world over worry about, nor the parish priest, who must address his "flock" mounted on a wooden (!) pulpit and who is but a small and severely constrained part of the Roman Catholic Church, would take kindly to an image, well-meant or no, stressing its precarious position within the cogs of a larger clockwork. Lastly, a cuckoo is held in high regard for its melodious staccato. A country girl would take no offense to being likened to "a little cuckoo" in a suitable context.

But do either Skinner's pigeons or Winston's Lispings deal with such disparate context-sensitivity? I can surmise that the pigeons don't. Winston could claim that to account for this phenomenon one only needs to increase the number of attribute/value pairs a frame posesses. This would be along the lines of Minsky's conjectures that intelligence is just keeping track of several hundred

special tricks. If so, then I would like to question that Marrian sentiment so piously hammered into the MIT AI-bound undergraduate, a notion explicitly embraced by Winston: An implementation is in order, an implementation able to replicate disparate context-sensitivity via transfer-framey static attribute/values. Until this is done, by Winston's own standards, no one is free to assume that static value/attribute pairs, symbol-pushing, and tree-walking are good AI ideas.

A burning question then arises: Why is there so much conviction in Winston's assessments? Why this certitude that good work is being done and that extensions of it will follow spontaneously and unproblematically? Well, [WINSTON 79] is but a tip of the iceberg crystalizing for over a decade now around [SIMON & NEWELL 73], a "fundamentalist" AI gospel that "guides" the world to embrace precisely these ideas. (*The production rule is my shepherd, I shall not want...*) But what has been wrought under the aegis of its weighty methodological orthodoxy? The outrageously funny old predictions of AI's success are ever replaced by equally firm, if less raptured visions, as unabashedly paraded in [ARTIFICIAL INTELLIGENCE 85]. Upon reflection, one does not find such certitudes amusing, if they have been stifling the field for decades with theoretical strangleholds akin to religious dogma. It is high time to kick the sleeping dog, so to speak; we must stop the twenty-year masquerade of inadequate formulations. If "learning through simile" is neither, it is neither proper nor innocuous to name it so. Doing so does not benefit AI's future or shore up its credibility, and neither does claiming that today's impotent techniques can be extended with a few hundred thousand more D.O.D. dollars and a consenting graduate student or two.

The bottom line is that you get what you think for. You think for a little bit and you think up *Robbie is like a fox* and proceed to build your S-expression castles on it, and so you get transfer-frames with conjunctive hierarchies of frames, every frame chock-full of simple attribute/value Lisp atoms. This does

not bring much color to the Polish cuckoo, not even to the English one, especially if one of your goals is "the competence" of curiosity. [WINSTON 79, p. 353].

Winston's theory-formative failings are common to the field, and are likely AI's most notorious deficiencies. One would hope that we could counter our natural tendency to hang on to our old ideas through the employment of anticipatory metaphorizing for the task of stress-testing and adjusting our thinking/ approach/model/solution. Such anticipation is what continues to make [TURING 50] such an enduring piece of writing. Incidentally, Turing and his work are absent from most general AI textbooks, including [WINSTON 84a]. Too bad.

III. The Case of a New Science: Lenat's Heuretics

Douglas B. Lenat's EURISKO work is not even tangentially the subject of this section, but his attendant discussion of Heuretics is. Heuretics, as [LENAT 82] points out, was proposed as a field of study by G. Polya in the 1940s. The object of Polya's study was heuristics: their origin, source of power, and how to acquire them. Lenat, apparently a mathematician by training, takes up Polya's idea, claiming to have a clear line of attack on the issue. Lenat initially defines a heuristic as "a piece of knowledge capable of suggesting plausible actions to follow or implausible ones to avoid."

As motivation for a present-day Heuretics, Lenat cites "the impressive performance of expert knowledge-based systems", and the attendant difficulties in getting humans to cogently convey their expertise to the builders of such programs. Therefore, one might as well build a program to synthesize expertise by itself. As methodology for pursuing Heuretics, Lenat demands "an accepted experimental procedure", and consequently suggests the building of programs which use heuristics and which try to find new ones.

At this point, I would like to convey my excitement and admiration for the theses stated above. I find it positively fascinating that Lenat has the gumption to give up on the human experts and artificially produce wisdom, and I find it refreshing that, true to the scientific tradition of Occidental Thought, the methodology to be used in this quest will be "accepted" and "experimental", never mind all those insight-driven, querulous AI workers. Of course, Lenat's own AI results (of announcing the guidelines for achieving Heuretics and his meta-methodological contributions) are, regrettably, insight-driven, but I anticipate no more of it once the programming picks up.

Shortly thereafter, however, Lenat takes a turn that I am unwilling to take: Deeming his initial definition of heuristics insufficient because it fail to stipulate

that each heuristic is to have a context in which it is used (or enjoined from use), Lenat replaces the definition with a different one. In fact, immediately after voicing his complaint, he writes: "In other words, heuristics must have both an if- and then- part." Insight-driven, data-driven, or just driven, I cannot see for the life of me why a piece of knowledge capable of suggesting plausible actions to follow or implausible ones to avoid *and* which comes with a user's manual and a manufacturer's disclaimer must be —is "in other words" —an if-then rule.

To put Lenat's conclusion in perspective, a few examples of annotated heuristics follow. While these examples are meant to be most unusual, Lenat's own original definition of a heuristic was used to generate them:

Heuristics can be metaphors. Are metaphors rules? Some say that metaphors are analogies, and I have seen logicians analogize in logic, but theirs is an act of faith. Metaphors subsume heuristics. An example of a complex modality-spanning metaphor would be the process of remembering a Laurie Anderson lyric (*Your eyes, it is a day's work to just look into them*) upon encountering some compelling pair of eyes. There is absolutely zero evidence for the prevalent and tacit AI assumption that metaphor is something static: a data structure or at most a mapping to and from data structures. In general, there is zero evidence for assuming that anything cognitive is static: a piece of Lisp on disk. My example metaphor is certainly nonstatic; call it a process if you itch for a definition. This metaphor is a nonlinguistic entity that includes the reminding as well as the embedded metaphor/trigger/quote. Could it be expressed as a logical proposition or combination of such? One can certainly write pages of logical connectives, *ad hoc* modal operators (ad hoc used as an AI smear), and oodles and oodles of scoped, parenthesized, and quantified objects in quixotic (and nowadays fashionable) attempts at devising one of those Barwise-and-Perryesque "syntactical expressions". These are used by New Wave logicians in their quest for the

representational Holy Grail, i.e., a propositional formalizaton of what happens in the head upon recalling Anderson's poetry while compelled by a pair of eyes [BARWISE & PERRY 83]. But avant-garde logic will not do. It hardly suffices for *George believes that Mary is clever and beautiful* [BARNDEN 85], a logic *wff* in sheepskin — as in [sheepish(George)] — that attempts to pass itself off for a member of the natural language family.

Clearly, my sample *metaphor-as-process* is a heuristic: It is a piece of knowledge to empathize with, to use, and to act upon. But, in addition to not being declarative, it is neither verifiable (save for empathy), nor presently measurable (we're waiting for Minsky-predicted future NMR and PET scanners to glimpse the patterns of thought —but of what use to AI would that be anyway?), nor formalizable through logic, nor statically expressible (save as a semiotic artifact or record thereof —but that begs the qestion of formalizing!).

What then is this *weird* heuristic? Let me present three AI metaphors for it. (1) It is an activity involving Freudian censors, pulsed through a hierarchy of frames. It is a collage of several thousand set pieces ("tricks") strung together for a fraction of a second by the software in charge of our senses and sensibilities [MINSKY]. (2) It is a massive activity of statistically describable, thermal motion-like message-passing and emergent symbol synthesis: a guideless, spontaneous, and thoroughly bottom-up free-for-all [HOFSTADTER 83]. (3) It is a snipet of a haunting image (haunting, because an aesthetical resonance is set-up in one's emotional context). This image is being evoked *involuntarily* through the process of associative memory recall. The evoking is accomplished by the metaphorizing cybernetics of the mind that are not too different from those of a carrot, actually, but we computationally know next to nothing of them. We will continue to know next to nothing of them until the simulation of emotional contexts as processes becomes fashionable to study, i.e., until it becomes an adver-

tised topic in AI conference announcements [my own three cents]. Notice that none of these three metaphors represents an if-then rule.

Heuristics can be images. Are images rules? Witness Romeo Void's *You cast a shadow but you don't stand up to the light!* Romeo Void's lyric is a heuristic because it allows us to visualize something that does cast a shadow but yet does not bear light well. We may perceive a light cone of a movie in progress, a photosensitive plate, a holographic image, or maybe something transparent, or delicate. The lyric also evokes a second visualization. From it we perceive a person who is self-important but who does not measure up in some important way. What's more, because it is an image, any verbalization or an attempt to *syntactically freeze* it does not stand up to the light! The image is a teeming process that creates tension between the two perceptions. It is a *metabolism* (!) that casts a shadow of meaning but which somehow cannot be made satisfactorily explicit using typography. Mere attempts at formalizing it *on paper*, i.e., as a data-structure or a *piece of logic* involving fixed entities obstructs it and makes it go away. There can also be many other readings of this lyric, but that is usual for image sources, and merely represents a yet another nail in the coffin of the notion that heuristics, and by close association, images, are if-then rules.

Heuristics can be memories. Are memories rules? A particularly vivid memory can prototypically affect one's learning for life. I don't think that we are in a position, Lenat included, to call memories if-then rules. It would be somewhat embarrassing to stake such a claim at a time when we cannot even simulate well the process of human remembering, regardless of mechanism. Our shoving bits into memory locations is remembering only as much as computer memory's page-trashing is the process of being anxious. There is, however, a lot to be gained from original new insights in distributed memory work, such as [KANERVA 84].

Heuristics can be smells, sights and sounds, and I don't mean the remembered human perceptions, like human olfactory memory (which come under the previous paragraph's scope). Are smells, sights, and sounds rules? Devices and decision procedures often use smells, sights and sounds —as opposed to abstractions — and their actions can be perceived in terms of if-then rules readily enough. Under devices and decision procedures, one would group here the homing instinct of pigeons, the human sexual response, the multichannel analyzer's display, or the Charge Coupled Devices atop Palomar Mountain, collecting the scant photons of a distant sight. The various phenomena — to invoke a dangerous, overworked word for smells, sights and sounds — are heuristics, for they constitute pieces of knowledge capable of suggesting some plausible actions to follow. Some reader-representationalist will jump all over me, demanding that only the cognitive agent's representations of phenomena may aspire to the status of heuristics. This view, however, is peculiar to relatively modern mentalists, and would certainly need to be defended against competing views from all over, ranging from that of the behaviorists to those of the quantumchromodynamists. Personally, I see no point in committing oneself to any view, save for this one: *Phenomena + Devices = Representations* (with apologies to the structured and logic-programming ayatollahs Dijkstra, Kowalski & Wirth). Either way, phenomena are not if-then rules, at least not all phenomena (!).

Governments are heuristics. Are governments rules? Governments came into existence when societies evolve mechanisms for conducting the business of "social intercourse", to borrow a phrase from John Donne. Governments are heuristics because they approximate solutions to existing problems, as well as embody all kinds of never tested *a priori* devices and decision procedures for handling unanticipated problems. It is precisely the dire straits of a government unable to fulfill society's expectations that channel both the true believers' holy

wars and the ad hoc committeeing. Is Lenat ready to say that governments are if-then rules? Again, while I concede that much of what governments do can be cast in terms of if-then rules, the governments themselves remain unruly (!).

The behavior of a mob, or any aggregate, can sometimes be described using descriptors which look like rules. But are mobs (aggregates) themselves rules? The behavior/metabolism/ biochemistry interface in humans can be told, given enough effort and insight, via rule-like descriptors, or so implies the attitude of the workers in the pertinent field [PSYCHONEUROIMMUNOLOGY 85]. But are people rules? Similarly, is mathematics a rule? Is everything a rule? If all these things are rules, are they all if-then rules? Are some merely if- rules or just then-rules?

Lenat inherits productions in people from a Herbert Simon and Alan Newell model of cognition ([SIMON & NEWELL 73]), formal implication from logic ([SMULLYAN 68]), and condition-action behavior from expert-systems programming ([DAVIS 82]). But whence is his knack for glossing over non-Positivism from the Neanderthals to contemporary European philosophy and the Australian aborigine? With heuristics, he wraps all of experience into interpreter-driven if-then rules. How formal of him. On the other hand, he introduces [LENAT 82] with the words: "heuristics, or informal judgmental rules". Why informal now? We just heard that heuristics have if- and -then parts? And the parts have slots, and the slots have values. They are data structures! What's an informal rule or an informal data structure, anyway? Is it like Stanislaw Lem's "simulated song" ([HOFSTADTER & DENNETT 81], [LEM 65])?

The problem with Lenat is that he uses his considerable fondness of mathematics to amplify and compound the sins of mathematicians from way before Gödel. His metaphors could have been gotten post and parcel out of Enlightenment magna opera (minus notation and theology) as they convey so

well the air of An Explanation of All There Is. Lenat not so much sleeps the AI's "Boolean dream" ([HOFSTADTER 85]), dreamed by many an AI worker in a fitful pursuit of a formal calculus of intelligence. Rather, he is snoring an Age of Reason tune in this Age of Information. When expounding on his ontology of heuristics, in [LENAT 83], he lights up an algebraic flare: "APPROPRIATENESS(Action, Situation)" (Shall we compute the derivatives and calculate its curl and divergence?), "non-linearity of the Situation space" (Not Hilbert —but is it a Banach one?), "the graph of heuristic's power as a function of the task it is applied to" (Gee, is it a mapping into reals? And what is the domain metric?). He even quotes Poincare's nineteenth century optimism, proving thus once and for all that the mathematicalizing metality is continuous and spans centuries: "The whole idea of analogy is that 'Effects', viewed as a function of situation, is a continuous function". Given Poincare, I feel induced to echo with my own favorite mathematical analogizers, The Waitresses: "as a function of variations on the curve, well that can be observed!" ([WAITRESSES]). Granted, The Waitresses mean girlwatching (alas, merely applied math), but as Hofstadter once exclaimed upon finding (some) of the meaning of life in the uncanny emergence of cooperation in the iterated pairwise Prisoner's Dilemma: "How rife this is with morals...!" [HOFSTADTER 85]. How well put.

With all this formal metaphorizing, one would think the object of the discourse was modern algebra or analysis, not fuzzy, fuzzy, fuzziness. Nay, I take it back. In real mathematics assertions get proved. Here, they merely get asserted —in mathematico-lingo, replete with corollaries and graphs that come on the heels of implicit counterfactuals such as "if we could somehow graph the...", before being unabashedly graphed, and then examined for inference of properties. If there ever was a want for examples of bootstrapping, [LENAT 83] would do fine: "Notice the area under the curve appears roughly constant", we are urged when

gazing over a nonquantified abscissa. What are are we to infer, that the implied scale is linear, not logarithmic?

Funny thing, Lenat's own paper fits his definition of a heuristic, except that its owner's manual lies (usual in the computer trade) and its manufacturer's disclaimers are curiously muted and often later implicitly counterclaimed. But, by Euclid, wherefore art thou, the if-part slot-label?

To be fair to Lenat, he goes on to say that:

> The whole process of drawing the power curves of heuristics is still conjectural. While a few such graphs have been sketched, there is no algorithm for plotting them, no library of thousands catalogued and plotted heuristics, not even any agreement on what the various power and task axes should be. Nevertheless, it has already proven to be a useful metaphor, and has suggested some important properties of heuristics. [LENAT 83]

This passage took the word *metaphor* out of my mouth. It is good that Lenat is aware, in the final tally, that drawing his curves is a metaphor, good or bad, a metaphor like any others. Why does he then seemingly refuse to see that the if-then rule is a metaphor as well, not some Kantian, ontologically valid noumenon? Why does he not say that, in fact, the sheer concept of a heuristic (yes, one can see through it!) is but a metaphor, in fact, a metaphor for metaphors. (Translation: "the class of metaphors", you know, the one in the Frankel-Zermelo Universe).

Lenat does no one a favor when seizing upon the chance to occlude the obvious (the reasons for expert systems' perceived successes) with the obfuscating (the call for a formalized Heuretics study; the promise of the if-then rule as the Alpha and Omega of computation, artificial or DNA). His knowledge-engineering cheer is sadly reminiscent of the feeling of power a kid gets after learning to multiply by hand, then discovering a calculator. Neither is the calculator doing anything awesome, nor are the EMYCINs of Palo Alto and points East. Each one performs an easily formalizable process, and both likely have more in common than

Lenat's heuristics and Lenat's brain's own metaphors. Although the
AM/EURISKO-gained insights and Lenat's considerable achievements as program-
mer are among the most interesting and original in AI, he can be expected to give
us far more if he could only successfully hack his one-sided and, as a result (unin-
tended, let's hope) deceptive metaphorizing.

IV. The Case of Emotions, Human and Computer: Turkle's Second Self

Sherry Turkle's new book is a joy to read. It provides fascinating glimpses of people's notions of self under revision because of their contact with computers. It also reports on the subjectivity inherent in gender (!) ([KELLER 85]) and its importance to how people use computers. Surprisingly, in view of Turkle's concern for male biases and predispositions, what has to be questioned in [TURKLE 84] is her unawareness of her own assumptions about computation, emotion, and intelligence that pervade her research from data acquisition to making prognoses.

Turkle does not directly concern herself with creating programs, but she does engage in model-building and theory-formation identical in character to that of typical AI research. Her unawareness of her own metaphorizing is a sobering example of what all AI researchers must strive to avoid.

Of reason and passion, Turkle writes this way:

> In Greek times, the experience of the divided human was captured in Plato's image of a driver of two horses, white and black, reason and passion. The ride was uneven. With or without the computer, people have found ways to describe this uneven ride. They use different languages: there are reason and passion, logic and emotion, ego and id. The computer makes a new contribution. Along with a new urgency, it provides a new discourse for describing the divided self. On one side is placed what is simulable; on the other, that which cannot be simulated.
>
> We cede to the computer the power of reason, but at the same time, in defense, our sense of identity becomes increasingly focused on the soul and the spirit in the human machine.
>
> Where we once were rational animals, now we are feeling computers, emotional machines. But we really have no way to put these terms together. [TURKLE 84, pp. 312-313]

But we do. Minsky does so in a couple of clear and concise words in [SCI-ENCE85] to the effect: How unfortunate it is that the old Platonic dichotomy between reason and affect keeps on being propagated through our time, despite evidence to the contrary; after all, people can rationalize emotional responses but cannot account for having thoughts.

Turkle is a respected scholar, a certified clinician, and an MIT professor of

psychology and sociology. Her sharp intellect and sensitivity are usually in plentiful evidence. Thus it is most discouraging to have her be unaware of the fact that she is propagating an a priori distinction set up by an ancient thinker without the benefit of, for example, our knowledge of the affective psychosis and its effect on judgment. She unconsciously propagates this reason/affect *apartheid* in her own modern assessment of the human and the artificial, with hardly a mention of the likely alternative. Yet hers is a book which shows that she had ample opportunity to handle and dispassionately convey the thoughts of other persuasive thinkers, from Hofstadter and Minsky to John Searle and Joseph Weizenbaum. Somehow her powers of critique have failed her when it came to detecting, isolating (or even suspecting) her own presuppositions, the ones that were *hard-wired* during her graduate youth, years before taking on this project.

Turkle is a scientist whose training is supposed to emphasize the very subjectivity of the conclusions drawn from observation. She even airs this issue from the angle of sexual development [TURKLE 84, Ch. 3]. It saddens me that her interests on this score stop at the politically attractive gender issue, well short of the matter of her own scholarly upbringing. This is ironic in light of the obvious care she has taken in experiment design and method: She takes up many pages to relate to the reader her mind-set and data gathering. Her Harvard education, heavy with Freud, has seemingly immunized her well from the neurobiochemical traditions of MIT's empiricism, even though she strongly credits her book's existence to the MIT environment.

MIT or Harvard, psychiatrists see emotion and thinking as inseparable. Every year our understanding of how neuropsychopharmacology explains emotion becomes more detailed. For example, the long suspected role of the brain's norepinephrine receptor sites in the pathology of affective psychosis [SHOPSIN 79] was recently demonstrated: the presence of lithium ions has been found to regu-

late the density of these sites in accordance to a predicting hypothesis based on oberved lithium efficacy in treatment. Yet, while lithium was first prescribed in Australia in 1946 for the treatment of depression, the entrenched psychotherapists in the American Psychiatry Association did not permit its use in America until the 1970s. It is now generally conceded that lithium, about the only massive success story in the history of psychiatry, gives normalcy to thousands of people whom psychoanalysis never did any good [WOLPERT 77]. Still, the psychoanalyzing Freudians remain unconvinced, shunning the molecular explanation for the cause of mania and depression in favor of their protracted touchy-feely explorations of the unconscious, even though their prophet wished so keenly for the emergence of "an organic basis" for his theories ([FREUD]).

As a practicing clinician, whatever her professional leaning, Turkle doubtless has witnessed patients with affective disorders. She has seen their impaired judgment when manic or depressed reverting to normalcy when they were returned to their euthalmic (normal) state either by medication or by the cyclic nature of the syndrome. (The work of the latter often wrongly self-attributed by many psychotherapists.)

Turkle knows that a great mathematician will not function as one when affectively psychotic. [FIEVE 75] and [DAUBEN 79] offer glimpses of how the minds of some great people —Winston Churchill, Abraham Lincoln, Geörg Cantor — would become incapacitated when their moods turned acutely high or low. These people would return to full bloom once their mood leveled off close to normal. For example, Cantor's life shows the ebb of his mathematical prowess (and the amplification of his religious convictions of the divine nature of transfinite numbers) with the recurrence of his mood swings, for which there was no effective treatment in Halle University's 19th century Nervklinik.

Incidentally, hypomania actually amplifies the cognitive powers incumbent

in the manic individual. Some great deeds of history and countless deeds of everyday *workaholism* or *drive* are the result of elevated mood. Mild depressions, too, have been credited with enhancing the aesthetics of such perennial sufferers of the blues as Emily Dickinson, Henry James, and Vincent Van Gogh. How is it then possible to dissect human mental activity into reason and passion? Emotion and cognition are one. And one need not appeal to the manic-depressive psychosis to find evidence for this. Affective disorders do, however, offer the AI researcher some still unexplored avenues for deep insight.

Yet, with all this known to her, Turkle lets her Freudian allegiances as psychoanalyst and Platonic sensibilities as a Freudian take the better of her deductive and inductive faculties: She is oblivious to her own use of the *thinking-as-reasoning* and the *emotion-as-non-computation* metaphors. She is not aware that these metaphors restrict her view of emotion, and therefore, her interpretation of people's self-perceptions. She does not see emotion as yet another observable — nothing sacred or mystical — not any more mystical than the ability to quickly recognize one's grandmother, which is plenty mystical to us all in today's AI ([HOFSTADTER 85]).

What is known about people and computers gives no reason to suppose that computers will never *have feelings*. If these feelings are to be seen as simulated, they will be no more simulated than the simulated thinking now conceded to EURISKO. When this becomes possible, the distinction between simulated and real will matter only to Dreyfus, Searle & Company. That ghost in the machine (emotion) will have a similar effect on those perceiving it as EURISKO's *simulated thinking* about three-dimensional VLSI design does now. Please note that I am not judging here the current programming achievements, whether they should be called *thinking, simulated thinking, mechanized inference, reasoning,* or *constraint-satisfaction*. I am merely asserting that the analogous programming involving

emotion can be had if we attempt it. What is *hot* in AI depends on what tickles those in the position to do the work. Already, Wendy Lehnert, Minsky, Schank and many others are beginning to explore some computational ideas on emotion. Surely it would be ill-advised to regard as intractable the setting up of an emotional context in artificial machines. In fact, the presupposition that machines and emotion are somehow apart (and the orthogonal and muddled row over consciousness), begat the *emotion-is-arcane-and-difficult-to-model* mind set.

But once AI recasts its world view into less Platonic terms (bluntly: once the old guard dies off), the AI problems and results are bound to take a different tack from the one that is responsible for our current backwater drifting. Hacking emotion into computers is contingent on the technological maturing of the AI community, not AI equipment. This maturing has nothing to do with the silly hype of the Fifth Generation Computing Project ([FEIGENBAUM & MCCORDUCK 83]) or the current expert system goldrush ([WINSTON 84b]). It it instead contingent upon the rise in sophistication of metaphor use by the people best equipped to study *process* and *computation*: the AI workers. It has equally little to do with any inherent problem intractability, or with the practical or theoretical limits of computer power, or with whether one is lucky enough to have a Symbolics in one's office (or a MacIntosh, depending on aspirations).

The very fact that most AI work to date has concerned itself with the modeling of thought as logical reasoning stems from the lack of understanding of emotion and its role in cognition and from the mathematicians, friends of reason from day one, having invented the toys with logic in mind.

Also, as covered by Turkle, there seems to be evidence that having a particular gender influences what one studies, how one studies it, and what one finds out about it ([KELLER 85]). How ironic it is that an ancient Greek's guesses should haunt us so far into the future! Is it because these guesses seem compel-

ling for societal reasons, i.e., emotion as regarded by males? Can anyone think of an equally enduring, pseudo-scientific notion lasting as long as this dichotomy of Plato has? Wouldn't it be eye-opening if the reason for its longevity was solely the science's *old-boy club* character? If so, then Barbara McClintock's 1950s work in cytogenetics and her fifty-years-late-in-coming male appreciation (Nobel Prize), as described in [KELLER 83], is writing on the wall for AI.

By conceding to humans the sole ability to feel, Turkle unwittingly distorts her data (the numerous conversations with people about computers) and impairs her own judgment. She forfeits the opportunity to formulate other questions which conceivably might have gained her (and her subjects) a far different understanding. True to Heisenberg's Uncertainty Principle, she disturbs the state of the observable with her observations, yet hers is an impaired judgment not mandated by any known constraints. It is entirely avoidable through caution.

It is instructive to AI people everywhere to see this confluence of impaired judgments: Plato's, when perceiving the mental; the experimental physicist's, when perceiving quantum phenomena; Turkle's own, when perceiving the forever human; the manic-depressive's, when. . . perceiving. To be aware of one's metaphorizing is to win half of the theory-formative battle, whatever the battle.

V. The Case of Slipability, Semanticity, and Musing: Hofstadter's Copycat

Douglas R. Hofstadter is not only an AI researcher but also a prize-winning author and a consummate writer. His metaphors for computation are, to my taste and knowledge, the best in the field. I thought it would be fitting to close this paper with a look at some excellent metaphorizing, as well as some metaphorizing excesses that can be seen to occlude Hofstadter's reporting, and possibly his research activity itself. The problem lies squarely with recognizing when too many metaphors are applied to the task of describing one's system, and when these metaphors become an illusion of comprehension deceiving the author himself. In the final computational tally, all recursion better *bottom out*, and so is the case with the AI metaphor: Sooner or later one must face implementing it. And I vote for *real soon*.

[HOFSTADTER 84] is a study of analogy-making in an impeccably designed domain, setup to probe in the abstract the sort of analogical phenomena that underlie metaphor, thought, recognizing one's grandmother instantaneously —the whole kit and caboodle, it seems. On the surface, with all its *if ABC goes to ABD, what does XYZ go to* stuff, it is more reminiscent of a game than of a project with implications for the knowledge of human thinking and perception. This may help explain the mystery of why it is so difficult to get funding for this sort of work, now that we have researchers deftly analyzing *Macbeth* and *Hamlet* for similarity and analogy, with a fistful of *conses*. . .

The notions underlying Hofstadter's approach are eruditely presented and compared to many others. The bulk of the report (without the appendices) is aptly titled "Methodology", and the reader is taken on a lucid journey through the project's intended competences, starting with, at first blush, a set of likely solutions to a trivial analogy, and culminating in a step-by-step analysis of a hypothetical solution to a truly difficult one (in fact, the one quoted above).

I would like to address [HOFSTADTER 84]'s spectacular successes and its one remarkable failing. The successes spring from the apt metaphorizing; the downfall becomes obvious only when attempting to actually implement the ideas: Not enough is specified to permit a reasonably full implementation.

The successes: Hofstadter models his computation on a biological cell. This is a powerful and uncanny AI view. He extends it with the same intensity as Lenat does his Heuretics, the difference being that Hofstadter's inventiveness of language and lucidity of description surpass Lenat's pseudo-mathematics. Hofstadter's description is also purely functional and never definitional. He never discusses the ontology of fuzzy objects, he merely creates computational creatures that *act like* enzymes on other creatures. Hofstadter uses adjectival and visual metaphors to great benefit. He talks of his diminutive procedures as "codelets", and of their repository as a "coderack". The codelets that are to explore as opposed to change things permanently he calls "musing". His spreading activation semantic network is a "Slipnet", for it allows the "fluid descriptions" to "slip" from one configuration to another, their global pattern forming "flickering clusters" that are as fleeting, impermanent and reconfigurable as the physics' flickering clusters being analogized. When giving reasons for the random execution of his "codelets", he uses the metaphor of a biological cell:

> In a cell, all activity is carried by enzymes. Enzymes of various sorts are distributed at random throughout the cytoplasm (the cell's interior), and because of random motion taking place inside the cytoplasm, they encounter all sorts of molecules in a very short time. Each enzyme has one or more (usually two) active sites. . . When an enzyme encounters a molecule that fits one of its active sites, it latches onto that molecule and fills that site. When all of its active sites are filled, the enzyme then performs its function, which may be constructive (combining two substrates into a larger molecule), reconstructive (changing the structure of a substrate), or destructive (reducing substrates into their components on a lower level of molecular structure). Usually, one enzyme's action is but a small link in a long *chain* of enzymatic actions whose collective result is the buildup of some complex product. . . It is to be emphasized that the cell relies on the random peregrinations of molecules inside it for these activities to be carried out: there is no Director General who observes all and shunts all pieces to their proper places at the proper times. This is simply because a cell is at too low a biological level to have such a centralized intelligent agent. A cell's 'intelligence', such as it is (and it is astonishing!), must emerge from the interplay of thousands of small,

independent processes whose outcomes have effects on the further activities to take place. A cell's 'intelligence' is of necessity highly distributed, in short, with waves after waves of enzymatic activity creating order out of chaos. In particular, the products made by one set of enzymes become the substrates to another set of enzymes. One remarkable feature of the cell is that enzymes themselves are produced, altered, and destroyed by other enzymes, so that the enzyme population of a cell is incredibly sensitive to the 'needs' of a cell: it is constantly adjusting itself according to the types of substrates present and absent. Elaborate feedback loops regulate the enzyme population. This rich metaphor has repeatedly furnished ideas for the design of the Jumbo and Copycat projects. [HOFSTADTER 84, p. 12]

This metaphor Hofstadter calls "subcomputation", and [HOFSTADTER 85] contains an eloquent argument for it as the way to do AI. His own AI ideas, based on non-"Director General"ed computation, are mirrored in the programming language research of Carl Hewitt and his MIT Message Passing Semantics group [HEWITT].

The cell metaphor, and the explicit step-by-step hypothesized solution of a hard problem are the shining examples of excellence in [HOFSTADTER 84]. They are also responsible for its ultimate failure as an AI technical report. Charmed by the metaphor, Hofstadter ultimately forgets to design the actual algorithms needed to implement his computational theory, and his step-by-step summary is on too high a description level to be of help. What happens is that Hofstadter writing the report and the reader perusing it are captives of the aggregated metaphors and simply do not notice just how much computing remains to be done. After reading [HOFSTADTER 84], one gets the urge to rush out to the nearest Lisp Machine and write it all up in code and watch it hum. And, yes, it can be easily done. . . all the way until we reach Step 3, "The Rule Generation Phase". Having just written 64K characters of code, we casually turn the page and read as follows:

The crucial question now is: How does a *rule* get manufactured, describing the visible action in a plausibly intensional way? After all, this is where we have been headed all along: towards a reasonable perception not only of X1 and X1*, but also of their interconnections, so that finally a good description of the change will 'fall out'. . .

To each concept in the Slipnet there is attached a measure of *semanticity*. This is a pure number, empirically assigned, which tries to estimate how useful a given concept

is in making characterizations of an object. . .

Exact numerical semanticity values are yet to be assigned, and certainly changing them would dramatically affect the performance of the Copycat program. This will be a most interesting and undoubtedly critical experiment to perform, since different settings of these crucial numbers surely would result in very distinct and characteristic cognitive styles. [HOFSTADTER 84, p. 24]

Well, I'd say that calling an infinite loop or a thrashing do-nothing cell a "cognitive style" is some metaphorizing! In fact, it is precisely the wrong kind, for it seduces everybody —Hofstadter included —into thinking that the homework of precisely describing and comprehending the system to be implemented has been scrupulously performed. Has either Hofstadter or anyone else —I certainly tried and failed —ever gotten beyond Step 2, "Moving into the Semantic Phase"? Perhaps we are not moving as much as we think we are? This leads me to question, of all people, David Marr. Was Marr on the right track advocating a clean break between implementation, algorithm, and points North? Perhaps the competence/computational-theory/algorithm/implementation strata are a web of interconnections, and cutting them, like lobotomizing a brain?

Corollary (apologies to Doug Lenat): Perhaps Marr's recommendations are like the nineteenth century surgical thought for the nineteenth century psychiatry? Perhaps the correct thing to do is not to cut but to be aware of connections and to attempt microsurgery only when all else fails, treating instead with one's metaphors —the new, synthesized neuroleptics. (And if that fails, there is always the electroshock and the behavior-modification therapies of peer review!)

As much as Winston's thinking is colored by his old code, specifically, by his ARCH system, so is Hofstadter's tinted by his cell metaphors. As much as Hofstadter is ahead of Winston in the development of AI conceptualizations (metaphor use, mostly), he commits just as grave a sin, that of not specifying the requisite computation enough. Hofstadter is a hardnosed AI researcher whose views often put him at odds with the AI establishment, which is ironic since the harshest AI

critic there is, John Searle, mistakenly attributes to Hofstadter the role of AI's spokesman. Yet, whatever Hofstadter's differences with the other AI workers, he shares their conviction that AI is about programs and ideas, not just ideas. I am sure that he is not willing to be relegated to the ranks of quibbling philosophers and does not intend to provide illusions of computational advances any more than Winston intends to provide illusions of understanding.

The pervasive villain in both cases and elsewhere is the metaphor, the grist too often in short supply for our churning AI thoughts. Paying attention to its harvest without getting reaped is as hard a problem as any we aim to take on. We cannot hope to go the distance burdened with negligence.

VI. Summary and Conclusions

This paper discussed how metaphors available to AI researchers *determine* AI research. They guide not just the choice and content of their projects, but also the thinking by which abstractions and methodologies come into being.

Its thesis is that metaphor use in the AI community is critical to the progress of AI, because most AI research is highly introspective, or insight-driven. AI research does not depend on external data so much as on abstract conceptualizations obtained through metaphor use. AI workers do not perform experiments, notwithstanding the field's self-avowed insistence on structured methodologies and the *program-writing-as-hypothesis-testing* stance, typified by Marr. The AI conceptualizations are instead direct descendants of metaphors used by AI researchers during theory formation and application. Some of these descendants are themselves metaphors, not algorithms or working programs. It is being slowly recognized that the MIT insistence on working programs as passports to AI legitimacy has accomplished a lot of business for the Lisp Machine vendors and has stimulated the production of turgid dissertations, so far turginating (with apologies to Lewis Carroll) in [SMITH 82], all the same yielding a pitifully scant amount of good AI ideas.

One conjecture following from this thesis is that AI researchers ought to report not only on their studies but also describe and clarify the extent to which their *results* are driven and supported by the metaphors they use.

Another conjecture is that more people seriously interested in AI ought to spend time thinking of the metaphor (and other conceptualizing devices) and its role in the shaping of AI. Perhaps the preponderance of hastily set-up sentence-prototypes of the sort *Robbie is like a fox* that motivate a lot of to-date work calls into question the worth of the AI research conjured to handle them.

AI has a long way to go before meeting its goals, and one of the roads it must take is that of methodological self-improvement. All AI methodologies are intimately intertwined with theory formation, consisting in fact of experimentally and observationally unconstrained sorts of theory formation. Also, AI would like to mimic engineering, when it comes to methodology, but the bases of AI research are nothing so solid as physics, genetics or mathematics. Instead, the AI brand of engineering is based on metaphor use and creative imagery. These metaphor-use and imagery-dependent *foundations* of AI must be acknowledged more explicitly, and the field's pretensions to being a branch of applied mathematics or logic must be dispelled. The only links that AI has with mathematics are those of being its bastard child and of having ex-mathematicians encrusted in its top echelons. These connections are harming AI because devising methodologies by analogy with mathematics has proven itself to be inadequate for the attainment of AI's self-described objectives, be it "the weak" or "the strong" AI [SEARLE 84].

VII. Bibliography and Acknowledgments

This section is organized in three parts. The first lists the authors and books constituting the subject of this paper. The second acknowledges the secondary, but by no means incidental, sources. These secondary sources are annotated. . . in a somewhat idiosyncratic way. A few sources lack annotations because although I was aware of them and have been somewhat acquainted with their content, I was unable to get acquainted with them thoroughly. (It is just amazing how many *important* books seem to walk away from the Indiana University's libraries!) The third is a list of acknowledgments, of people and works. My acknowledgments appear with my bibliography section because I owe thanks to all who shaped my thinking about AI, especially the authors of the works criticized in this paper.

VII.1 Subject sources:

By Douglas R. Hofstadter:

[HOFSTADTER 84]
> "The Copycat Project", Massachusetts Institute of Technology's Artificial Intelligence Laboratory Memo #755, Cambridge, Massachusetts, January 1984.

[HOFSTADTER 85]
> *Metamagical Themas*, Basic Books, New York, 1985.

[HOFSTADTER 82-84]
> Personal communication, 1982 to 1984.

By Douglas B. Lenat:

[LENAT 82]
> "The Nature of Heuristics", *Artificial Intelligence*, volume 19, Number 2, North-Holland, Amsterdam, 1982.

[LENAT 83]

"The Role of Heuristics in Learning by Discovery: Three Case Studies", in *Machine Learning*, R.S. Michalski, J.G. Carbonell, T.M. Mitchell, editors, Tioga, Palo Alto, California, 1983.

By David Marr:

[MARR 82]
Vision, W. H. Freeman, San Francisco, 1982.

By Sherry Turkle:

[TURKLE 84]
The Second Self: Computers And the Human Spirit, Simon and Schuster, New York, 1984.

By Patrick H. Winston:

[WINSTON 79]
"Learning by Creating and Justifying Transfer Frames", in *Artificial Intelligence: An MIT Perspective*, P.H. Winston and Brown, editors. MIT Press, Cambridge, Massachusetts, 1979.

[WINSTON 80-82]

MIT AI Lab Memos #520 (Revised May, 1980), #632 (Revised November 1981), #678 (Revised September 1982), #679 (Revised January 1983).

[WINSTON 83a]

Lectures in the undergraduate AI course at MIT, Fall 1983.

[WINSTON 84a]

Artificial Intelligence, 2nd edition, Addison-Wesley, Reading, Massachusetts, 1984.

VII.2 Other sources:

[AMAREL]

Saul Amarel, "Artificial Intelligence and the Social Sciences: A Preliminary Report", Rutgers University, Lab for Computer Science Research, Technical Report #60, November 1984.
[A brief listing of some expert system work. Otherwise, has nothing to do with "social sciences" or implications of AI.]

[ANDERSON]

Laurie Anderson, *Big Science*, Warner Bros., Bsk 3674, 1982.

Laurie Anderson, *Mister Heartbreak*, Warner Bros., 25077-1, 1984.

Laurie Anderson, *United States Live*, Warner Bros., 25192-1, 1984.
[Laurie Anderson is an avant-garde musician whose metaphors and imagery are on a par with Kate Bush's, in other words, poignant, startling, and humorous. Her music, as well as video and stage performannce, is equally special.]

[ARTIFICIAL INTELLIGENCE 85]

"Artificial Intelligence —Where Are We?", D.G. Bobrow and P.J. Hayes, editors, *Artificial Intelligence*, Volume 25, Number 25, March 1985, North-Holland Press, Amsterdam, 1985.
[An eye-openner, where metaphor is scarcely mentioned (!) and where, as usual, Hubert Dreyfus comes up with the largest number of original, refreshing, and controversial insights.]

[BARNDEN 85]

John A. Barnden, "Attitudes and Imputations", unpublished manuscript, Indiana University Computer Science Department, Bloomington, Indiana, 1985.
[A critique of prior propositional attitudes approaches, esp. Barwise & Perry's and Creary's. Apparently none of this stuff really works, and not for the lack of trying!]

[BARWISE & PERRY 83]

Jon Barwise and John Perry, *Situations and Attitudes*, Bradford Books: MIT Press, Cambridge, Massachusetts, 1983.
[A typical stab at coercing logic into representing beliefs and opaque contexts, with mixed results on very simple examplars. A dead-end.]

[BATALI 83]

John Batali, "Computational Introspection", MIT AI Lab Memo #701, February 1983.
[A compare-and-contrast-and-critique of Weyrouch, Doyle and Brian Smith. Very readable and time-saving.]

[BLACK]

Max Black, *The Labyrinth of Language*, Frederick A. Praeger, New York, 1968.

[BUCK 84]

Ross Buck, *The Communication of Emotion*, Guilford Press, New York, 1984.
[Interesting thesis that human and animal communication is based on emotion. Interesting in light of language-heavy orientation of AI communication research. Important for those interested in modeling emotion computationally. Lots of human nonverbal communication material.]

[BUSH]

Kate Bush, *The Kick Inside*, EMI-America, SW-17003, 1977.

Kate Bush, *Lionheart*, EMI-America, SMAS-17008, 1978.

Kate Bush, *Never For Ever*, Harvest (Canada), STAO 6476, 1980.

Kate Bush, *The Dreaming*, EMI-America, ST-17084, 1982.

Kate Bush, *Hounds of Love*, EMI-America, 1985 (in press).
[Kate Bush is a British musical artist whose mixed metaphors are a high art. Below is a chronological listing of her work. Also see one of the acknowledgment notes below.]

[CARROLL]

Lewis Carroll, *The Annotated Alice in Wonderland*, Martin Gardner, editor and annotator.
[My second most favorite book. An inspiration for all language-crazy folks.]

[CODE 85]

Murray Code, *Order and Organism; Steps to a Whiteheadian Philosophy of Mathematics and the Natural Sciences*, State University of New York Press, Albany, 1985.
[A highly readable stab against objectivism. Questions the cognitive significance of logic and mathematics. "The proper business of philosophy is the search for adequacy of expression rather than for certainty or truth." (p. 97)]

[DAUBEN 79]

Joseph W. Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Chapter 12, Harvard University Press, Cambridge, Massachusetts, 1979.
[Wonderful exposition of Cantor's mathematics and a well-done biography. Lots of fascinating quotes and facts.]

[DAVIS 82]

Randall Davis, "Expert Systems: Where Are We? And Where Do We Go from Here?", MIT AI Lab Memo #665, June 1982.
[A beautifully illustrated guide to expert sytemhood by a hardboiled expert systems man. Great fun to read.]

[DAVIS 83]

R. Davis and H. Shrobe, Lectures in the Graduate AI Course at MIT, Fall 1983.
[You had to be there. One of the sources of "Ad hoccery versus systematicity".]

[DIJKSTRA]

Edsger Dijkstra, *A Discipline of Programming*, Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[DOWNS & STEA 73]

Roger M. Downs and David Stea, editors, *Image and Environment; Cognitive Mapping and Spatial Behavior*, Aldine Publishing, Chicago, 1973.
[Engaging analysis of spatial problem-solving, such as what to do if one finds oneself in the middle of Milan, Italy. Lots of cognitive mapping research, including important classic papers.]

[DREYFUS 79]

Hubert Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence*, second edition, Harper and Row, New York, 1979.

[FEIGENBAUM & MCCORDUCK 83]

Edward A. Feigenbaum and Pamela McCorduck, *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, Addison-Wesley, Reading, Massacusetts, 1984.
[Breezy and entertaining trash. Go out and buy shares in my expert system company.]

[FIEVE 75]

Robert Fieve, *Moodswing: The Third Revolution in Psychiatry*, Morrow, New York, 1975.
[The story of lithium carbonate told by one of its fans. lots of gripping cases of manic-depressives, including historical figures like Teddy Roosevelt.]

[FISHER 85]

Harwood Fisher, *Language and Logic in Personality and Society*, Columbia University Press, New York, 1985.
[A new book summarizing and discussing some old ideas: personality, objectification (see [KELLER 85]), language, society, and logic — combined. Replete with thought-provoking passages that have a direct and immediate bearing on what AI is and should be worrying about. Worth the aggravation of sifting through some psychoanalytic mumbo-jumbo.]

[FREUD]

Sigmund Freud, *An Outline of Psychoanalysis*, James Strachey, translator, W.W. Norton, New York, 1949.
[The bane of psychiatry, actually invented because of lack of "organic methods". If Freud were alive today, he would be at MIT, studying neurotransmitters!]

Sigmund Freud, *Totem And Taboo; Some Points of Agreement Between the Mental Lives of Savages and Neurotics*, James Strachey, translator, Norton, 1950.

[Unbearably flawed and ethnocentric treatment of the Australian aboriginees. All-around nonsense. Sort of like watching a horror film.]

Sigmund Freud, *Jokes and Their Relation to the Unconscious*, James Strachey, translator, revised by Angela Richards, Harmondsworth: Penguin, New York, 1976.
[A lot of bad jokes and some interesting points about them.]

Russell Jacoby, *Social Amnesia: A Critique of Contemporary Psychology from Adler to Laing*, Beacon Press, Boston, 1975.

J. Rickman, editor, *A General Selection from the Works of Sigmund Freud*, Doubleday Anchor Books, New York, 1957.
[Good reference; beats lugging umpteen volumes around.]

[GENTNER 83]

Dedre Gentner, "Structure-Mapping: A Theoretical Framework for Analogy", *Cognitive Science*, Volume 7, Number 2, 1983.
[An unconvincing try at computational metaphor on the basis of inner metaphor systematicity. Won't work for non-standard metaphors, mixed metaphors, or creative metaphors.]

[GENTNER 83b]

D. Gentner and A. Stevens, editors, *Mental Models*, Lawrence Erlbaum, Hillsdale, New Jersey, 1983.
[Everything you always wanted to know about naive physics and more. The title is misleading, save for the chapter on Micronesian navigation.]

[GUMPEL 85]

Lisotte Gumpel, *Metaphor Reexamined*, Indiana University Press, Bloomington, 1984, in the series *Advances in Semiotics*, Thomas A. Sebeok, general editor.
[See one of the acknowledgment notes below.]

[HAYES 77]

Patrick J. Hayes, "In Defence of Logic", *Proceedings of the Fifth IJCAI, Massachusetts Institute of Technology*, 1977.
[Very amusing and poignant stab at the semantic network club.]

[HAYES 83]

Patrick J. Hayes, "The Second Naive Physics Manifesto", University of Rochester's Cognitive Science Technical Report #10, October 1983.
[A clearheaded assessment of logic's constraints as such, but regrettably advocating carrying on as ususal.]

[HEWITT]

[There are numerous sources of information about the Actors, or the Message Passing Semantics, research at the MIT AI Lab and at the MIT LCS (The Laboratory for Computer Science), under the direction of Carl Hewitt. Quite a few overlap significantly. The sources cited below servicably span that collection.]

Carl Hewitt, "Viewing Control Structures as Patterns of Passing Messages", in *Artificial Intelligence: An MIT Perspective*, P.H. Winston and Brown, editors, MIT Press, 1979.

William D. Clinger, *Foundations of Actor Semantics*, a Ph.D. thesis, MIT AI Lab Technical Report #633, May 1981.

Carl Hewitt and Peter de Jong, "Open Systems", MIT AI Lab Memo #691, December 1982.

Daniel G. Theriault, *Issues in the Design and Implementation of ACT2*, an M.S. thesis, MIT AI Lab Technical Report #728, June 1983.

Carl Hewitt and Henry Lieberman, "Design Issues in Parallel Architectures for Artificial Intelligence", MIT AI Lab Memo #750, November 1983.

Carl Hewitt, article on the parallel asynchronous systems' non-equivalence with a Turing Machine and non-reproducability with logics. Byte, April 1985.

[HOFSTADTER 79]

Douglas R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books, New York, 1979.
[Great dialogs, and more symmetries and puns than can be gotten. A great way to widen one's horizons.]

[HOFSTADTER & DENNETT 81]

Douglas R. Hofstadter and Daniel Dennett, editors, *The Mind's I*, Basic Books, New York, 1981.
[Lots of provocative writings on consciousness, intelligence, and simulation. A thoughtful treatment of Searle's "Chinese" metaphor.]

[HOFSTADTER 83]

Douglas R. Hofstadter, "Who Shoves Whom Inside the Careenium?, or, What Is the Meaning of the Word 'I'?", Indiana University Computer Science Department's Technical Report #130, 1982.
[A great little dialogue on the composite nature of symbols. A nice exposition of Hofstadter's subcomputation metaphor.]

[KANERVA 84]

Pentti Kanerva, *Self-Propagating Search: A Unified Theory of Memory*, Ph.D. thesis,

Stanford University Department of Philosophy, Center for the Study of Language and Information Technical Report No. CSLI-84-7, Stanford, California, 1984.
[A highly-original model of human memory involving computing the Hamming distance of items stored as binary strings and retrieved through "evoking". Complete with mathematical analysis of the proposed model.]

[KELLER 83]

Evelyn Fox Keller, *A Feeling for the Organism: The Life and Work of Barbara McClintock*, Freeman, New York, 1983.
[A fascinating life story of a woman-scientist whose work is quite unlike that of any man.]

[KELLER 85]

Evelyn Keller, *Reflections on Gender and Science*, Yale University Press, New Haven, 1985.
[A compelling book on the biases indtoduced in science by the gender of the scientist. A careful historical account replete with disgusting quotes.]

[KOSSLYN 80]

Stephen M. Kosslyn, *Image and Mind*, Harvard University Press, Cambridge, 1980.
[A huge report on imagery, and visual and linguistic metaphor, including computational implementation.]

[LAKOFF & JOHNSON 80]

George Lakoff and Mark Johnson, *Metaphors We Live By*, The University of Chicago Press, Chicago, 1980.
[A most readable treatment of metaphor, meaning, and truth. This may be the best read in all of cognitive science!]

[LEM 65]

Stanislaw Lem, *The Cyberiad*, M. Kandel, translator, Hartcourt Brace Jovanovich, New York, 1985.
[My favorite book. The best theory of AI text around. As wonderful for AI workers as [CARROLL] is for mathematicians.]

[LINDSAY 73]

Robert K. Lindsay, "In Defense of Ad Hoc Systems", in *Computer Models of Thought and Language*, Roger C. Schank and Kenneth Mark Colby, editors, W.H. Freeman, San Francisco, 1973.
[An articulate position paper discounting general syntactical methods in favor of modeling AI inference on observed human quirks. Even though the argument is based on directed tree graphs only, it generalizes well to all sorts of AI ideas. Some discussion of imagery and memory maps.]

[LINSKY 83]

Leonard Linsky, *Oblique Contexts*, The University of Chicago Press, Chicago, 1983.

[MCCARTHY 80]

John McCarthy, "Circumscription —A Form of Nonmonotonic Reasoning", *Artificial Intelligence*, Volume 13, Numbers 1 and 2, North-Holland, Amsterdam, 1980.
[According to the author himself, this is his best work ever.]

[MINSKY]

Marvin Minsky, "Plain Talk about Neurodevelopmental Epistemology", MIT AI Lab Memo #430, June 1977.
[In the continuing absence of Minsky's *Society of Mind* (chronically forthcoming), this paper is the next best thing to read.]

Marvin Minsky, "K-Lines: A Theory of Memory", MIT AI Lab Memo #516, June 1979.
[A mechanism for making plausible inferences in Minsky's model of the mind.]

Marvin Minsky, Patten Foundation Lectures, Indiana University, 1985.
[You had to be there. Mostly on "Society of Mind" and jokes.]

Marvin Minsky, Personal communication on metaphor and "Society of Mind", 1983 to 1985.
[Marvin Minsky is the most fun person to talk to in all of Brookline!]

[MITCHELL]

Joni Mitchell, *Mingus*, Asylum 505.
[A tribute to late Charles Mingus. Filled with captured idioms of black urban America. Good jazz, to boot.]

[NEWELL 85]

Alan Newell, "Project SOAR" research, Carnegie-Mellon University, Department of Computer Science, 1985.

[NORMAN 81]

Donald A. Norman, "Categorization of Action Slips", *Psychological Review*, Volume 88, Number 1, 1981.
[How your spoonerisms tell us how your mind works.]

[NORMAN & RUMELHART 75]

Donald A. Norman, David E. Rumelhart, et al., *Explorations in Cognition*, W.H. Freeman,

San Francisco, 1975.
[UCSD-Psych's old position book on AI: semantics networks galore and too much parsing. Will be made obsolete by the upcoming Rumelhart & McClelland (1985) tome.]

[ORTONY 79]

Andrew Ortony, editor, *Metaphor and Thought*, Cambridge University Press, London, 1979. [Interesting polemics on the no-winner philosophical plane. Interesting ideas to guide designing one's own philosophical/computational stance on the issues of linguistic metaphor, specially chapters by Boyd, Rumelhart, Searle, and Pylyshyn.]

[PASTUSZKA 61]

Jozef Pastuszka, *Psychologia Ogólna*, Towarzystwo Naukowe Katolickiego Universytetu Lubelskiego, Lublin, Poland, 1961. [This is a good historical treatment of many aspects of psychology, and is invaluable as a record of evolving attitudes about emotion from Plato on. Well worth translating from Polish. Stosunkowo przyzwoicie opracowana, szczególnie w miejscach gdzie naukowe podejście nie zgadza sie dokladnie ze stanowiskiem kosciola. (Przyjemnie tak po Polsku zasunać w artykule naukowym wydanym po angielsku w U.S.A.!)]

[PILEJKO 76]

Konstanty Pilejko, *Prakseologia —Nauka o Sprawnym Dzialaniu*, Volume 226 in the series *Biblioteka Problemów*, Państwowe Wydawnictwo Naukowe, Warszawa (Warsaw), 1976.

[PIOTROWSKI 79]

Zygmunt Piotrowski, *Perceptanalysis*, Ex Libris, Philadelphia, 1979. [Rorschach revisited and exquisitely defended. Of interest to the computational issues involving imagery.]

[PLATO]

Plato, *The Republic*, Grube, translator. [For us, Plato splits the mind into reason and passion.]

Plato, *Philebus*. [Plato invents psychology.]

[POLYA]

G. Polya, *Mathematics and Plausible Reasoning*, Volume 1, Princeton University Press, 1973.

[PSYCHONEUROIMMUNOLOGY 85]

Steven Locke, Robert Ader, Hugo Besedovsky, Nicholas Hall, George Solomon, Terry Strom and Herbert Spector, editors. *Foundations of Psychoneuroimmunology*, Aldine Publishing Co., New York, 1985.
[Classic papers on the behavior/metabolism issues important to AI.]

[QUINE]

W.V. Quine, *Ontological Relativity and Other Essays*, Columbia University Press, New York, 1969.

[RICHARDSON 69]

Alan Richardson, *Mental Imagery*, Routledge and Kegan Paul, London, 1969.
[A historical survey of mental imagery investigations and an overview of contemporary (late 1960s) thinking. Chapter 6, "Conclusions and Speculations" is particularly interesting.]

[ROMEO VOID]

Romeo Void, *Instincts*, Columbia, BFC 39155, 1984.
[Chockful of images, usually directed toward an uspecified spouse-like entity; also some diffuse ones. Very good music passionately sung by Debora Iyall.]

[RORTY 80]

Amelie Oksenberg Rorty, editor, *Explaining Emotions*, University of California Press, 1980.
[A must for the emotion hacker. A large collection of recent papers stressing 'relevant empirical investigations'. Among other things, contains articles rejecting the widely held in AI view that beliefs and emotions can be expressed with propositions.]

[SCHANK]

Roger Schank and R. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum, Hillsdale, New Jersey, 1977.
[The notion of script is developed here. Skip the plans and goals.]

Roger Schank, *Dynamic Memory; A Theory of Reminding and Learning in Computers and People*, Cambridge University Press, 1982.
[A psychological theory of memory, its change and assimilation of new experiences, and reminding —all in MOPs and TOPs.]

[SCIENCE85]

Science85, cover stories on artificial intelligence in the March 1985 issue. Quotes by

-

Dreyfus, Hinton, Minsky, Schank, Wilensky, and Winograd. Science85, AAAS, Washington, 1985.
[Highly entertaining. Wis h they also had sections with Searle and Hofstadter.]

[SEARLE 69]

John Searle, *Speech Acts; An Essay in the Philosophy of Language*, Cambridge University Press, London, 1969.

[SEARLE 84]

John Searle, *Minds, Brains, and Science*, Harvard University Press, Cambridge, 1984.
[AI's most heeded detractor in full swing.]

[SEMIOTICS 85]

Robert E. Innis, editor, *Semiotics: An Introductory Anthology*, Indiana University Press, Bloomington, 1985, in the series *Advances in Semiotics*, Thomas A. Sebeok, general editor.
[See one of the acknowledgment notes below.]

[SHEEHAN 72]

Peter W. Sheehan, editor, *The Function and Nature of Imagery*, Academic Press, New York, 1972.
[A good collection of papers by notable anti-behaviorists.]

[SHOPSIN 79]

Baron Shopsin,*Manic Illness*, Raven Press, New York, 1979.
[The only definitive reference on mania. Hardcore neuro. Hardcore pharmo. Lots of sexy psychopharmacology.]

[SIMON & NEWELL 73]

Herbert A. Simon and Alan Newell, *Human Probem Solving*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973.
[The "Holy Book" of top-down, sequential, symbolic, deterministic computation. Something akin to *Principia Mathematica*.]

[SKINNER 80]

Robert Epstein, Robert Lanza, B.F. Skinner, "Symbolic Communication Between Two Pigeons (Columbia livia domestica)", in *Skinner for the Classroom; Selected Papers*, by B.F. Skinner, Robert Epstein, editor, Research Press, Champaign, Illinois, 1982. Originally published in *Science*, Volume 207, p. 543, AAAS, 1980.

[A spoof of "Symbolic communication between two chimpanzees (Pan troglodytes)". Witness Jack and Jill, "two moderately hungry adult pigeons" (Jill's a theoretical girl, too...). The rest is, ahem, symbolic. True story! Even has a sequel!]

R. Epstein, B.F. Skinner, "The spontaneous use of memoranda by pigeons", *Behaviour Analysis Letters*, Volume 1, p. 241, North-Holland, Amsterdam, 1981.
[The exploits of Jack ad Jill, continued. See Jack make not-so-mental notes and (listen up, Roger Schank!) remind himself, distractions or not. Ditto Jill.]

B.F. Skinner, "'Superstition' in a pigeon", in *Skinner for the Classroom; Selected Papers*, by B.F. Skinner, Robert Epstein, editor, Research Press, Champaign, Illinois, 1982. Originally published in *Journal of Experimental Psychology*, Volume 48, p. 168, 1948.
[Skinner wrote this one while chairing Indiana University's Psychology Department, probably to spite the local humanities establishment. At least that's what his postcript is about: He implies to have demonstrated a mechanism that would explain away a large chunk of the Hopi culture, "the superstition of a rain dance". I'm sure that the Hopi will be contingently reinforced, i.e., take offense.]

B.F. Skinner, "Pigeons in a Pelican", in *Skinner for the Classroom; Selected Papers*, by B.F. Skinner, Robert Epstein, editor, Research Press, Champaign, Illinois, 1982. Originally published in *American Psychologist*, Volume 15, p. 28, 1960.
[A "moby hack" even better than Oliver Smoot, Jr.'s Harvard Bridge ordeal. (In case you don't know, the Harvard Bridge near MIT has been determined to be some 364.4 smoots (and one ear) long. The markings are repainted every year.) Here, some pigeons were trained to guide missiles, and well, too; rejected by the non-guidance-minded Pentagon (on the eve of Hiroshima) due to lack of imagination (so what else is new). Perhaps this methodology can be now recommended to the "poor" computer science departments which barely missed the Strategic Defense Initiative gravy but would still like to develop some artificially intelligent killing systems just to keep a finger in D.O.D.'s pocket.]

[SMITH 82]

Brian C. Smith, *Reflection and Semantics in a Procedural Language*, Ph.D. thesis, Massachusetts Institute of Technology, MIT AI Lab Technical Report #272, January 1982.
[Peter Szolovits, thesis advisor, apparently hasn't read this one yet, which is not surprising. At almost 500 pages of muddleheaded roller-coster ride from philosophical inanities to bizarre Lisp constructions, this *Reflection* may just be the most opaque and nonelucidating document in the history of MIT. Maybe even a shoo-in for the worst thesis ever done outside Education. An interesting idea, better summarized by [BATALI 83] than even the author's apologetic follow-up pamphlet.]

[SMULLYAN 68]

Raymond M. Smullyan, *First Order Logic*, Springer-Verlag, New York, 1968.
[*The* textbook on logic.]

[SMULLYAN 73]

Raymond M. Smullyan, *The Tao Is Silent*, 1973.
[A wonderful sabotage of our Judeo-Christian values.]

[TURING 50]

Alan Turing, "Computing Machinery and Intelligence", The Mind, 1950. Reprinted in *Computers and Thought*, Feigenbaum and Feldman, editors, 1963.
[Profound and amusing. Turing test described and defended.]

[WAITRESSES]

The Waitresses, "Pussy Strut", *Wasn't Tomorrow Wonderful?*, PolyGram Records, PD-1-6346, 1982.
[You'll love them. Patti Donahue, lead vocal. Wonderful superposition of everyday images into surrealistic collages. Somewhat cheeky, too.]

[WEYHROUCH 80]

R. Weyhrouch, "Prolegomena to the Theory of Formal Reasoning", *Artificial Intelligence*, Volume 13, Numbers 1 and 2, North-Holland, Amsterdam, 1980.
[Very nice effort at mechanizing logic and just as irrelevant to AI's future.]

[WHITEHEAD 27]

Alfred North Whitehead, *Symbolism; Its Meaning and Effect*, Barbour-Page Lectures, University of Virginia, 1927, reprinted by Capricorn Books, New York, 1959.
[An exposition of symbolism's 'suggestiveness and emotional efficacy', perception of causality, and cybernetics of action, well before Wiener coined the term. Recommended not only for the logic-bound mental models (qualitative physics) hackers.]

[WINSTON 84b]

P. Winston and Karen Prendergast, editors, *The AI Business: The Commercial Uses of Artificial Intelligence*, MIT Press, 1984.
[Use this book to convince your alumni to give your department money.]

[WOLPERT 77]

Edward A. Wolpert, "A Holisitc Approach to Bipolar Depressive Illness", in *Manic-Depressive Illness*, E.A. Wolpert, editor, International Universities Press, New York, 1977.
[A more "balanced" treatment of psychotherapy vs. drug therapy than warranted. Included here for "equal time" and all that.]

VII.3 Acknowledgment notes:

Lisotte Gumpel's book [GUMPEL 85] and Robert Innis's collection [SEMIOTICS 85] have given me a glimpse of what the semioticians and literature scholars think

of metaphor these days. Of special influence was the treatment of a *Wizard of ID* cartoon by Gumpel, a cartoon with an allegedly taunting cuckoo clock. My objections to some of the points and interpretations in the discussion of this image gave me the idea of using *Robbie is like a cuckoo*. In Innis, of much interest is Umberto Eco's paper on the semantics of a metaphor. In general, I feel that something is to be gained from crossing AI with semiotics, and I am now in the process of investigating this interdisciplinary potential.

Of considerable help in the case study on Lenat was the recent monograph by Edwin Hutchins, James Hollan, and Donald Norman of NPRDC-UCSD Intelligent Systems Group, called "Direct Manipulation Interfaces". In this paper the authors mention a working system for graphical interaction with statistics bypassing the usual medium of algebraic manipulation and programming. Their insights induced me to contemplate challenging Douglas Lenat's "The Nature of Heuristics". The version of February 21, 1985 was kindly made available to me by the authors, for which I thank.

Of unusual character is the contribution of vocalist and poet Kate Bush's lyrics, which inspired a variation on her theme in the guise of the *log* mixed-metaphor appearing herein. Her special imagery has helped me touch the rich viscera of metaphor, including the much-misunderstood mixed sort, and by contrast perceive the degenerate miasma of metaphor-use in AI literature and scientific writing in general. I would like to credit Doug Alan, a.k.a. nessus@mit-eddie, for supplying the bibilographical information on Kate Bush and Laurie Anderson, and for his continuing efforts to gain them a wider appreciation among the readers of Usenet.

I would especially like to express my gratitude to my teachers, Professors John Barnden and Douglas R. Hofstadter. In addition to greatly influencing my scholarly upbringing, Douglas R. Hofstadter also took me with him to the MIT