Symbolic Demeaning:
The Loss of Meaning in Human-Computer Communication

William H. Sterner

Conceptual Foundations of Science and Computer Science
University of Chicago

DRAFT -- Not for citation without permission from the author.

Comments welcome: bill@midway.uchicago.edu

Contents

Frame Setting for an After-Post-Modern and Anti-reductive Approach to Human-Computer Interactions....... 1
Anthropomorphism and ELIZA..8
ELIZA and natural language..16
Descent into human complexity..26
Demeaning features of ELIZA-person dialog........................38
Anthropomorphism and anthropocentrism /technical deepening and symbolic demeaning..........................53
Steps toward a moral economy of design..............................63
Logic Provides Only an Objective Standard that Falls Far Short of Grasping Lived Meanings........................68
Bibliography...70
Appendix 1 - Weizenbaum article..74
Appendix 2 - KantPro™ selections..75

000420

Frame Setting for an After-Postmodern and Antireductive Approach to Human-Computer Interactions

> "I got my first glimpse of artificial intelligence on Feb. 10, 1996, at 4:45 p. m. EST, when in the first game of my match with Deep Blue, the computer nudged a pawn forward to a square where it could easily be captured. It was a wonderfully and extremely human move. … Later I discovered the truth. Deep Blue's computational powers were so great that it did in fact calculate every possible move all the way to the actual recovery of the pawn six moves later. The computer didn't view the pawn sacrifice as a sacrifice at all. So the question is, if the computer makes the same move that I would make for completely different reasons, has it made an "intelligent" move?" Gary Kasparov[1]

In the first game of his match with a computer, Gary Kasparov experiences a dramatic shift in his characterization of Deep Blue's mode of play from apprehending a "wonderfully and extremely human move" to realizing that the computer "did in fact calculate every possible move." My main project in this paper is to examine the character of human-computer interactions as typically involving two participants making moves "for completely different reasons." What results is not so much a discussion of the "intelligence" of computers as an analysis of the losses of meaning humans commonly experience in interaction with computers. Like Theophrastus, the ancient Greek philosophical botanist and student of ethical types, and I will concentrate on "bad characters" or pathologies of 'symbolic demeaning' that emerge from the lack of communicative parity between a human and a computer program in symbolic exchange.[2]

---

[1] Gary Kasparov in response to Deep Blue as a chess partner: "The Day that I sensed a new kind of intelligence," *Time Magazine*, March 25, 1996, p. 55.

[2] Also like Theophrastus, I full well intend to treat the "good characters" or interactive benefits of such exchanges in other places and different ways. It is very important for us to explore collectively what human virtues turn out to be in a community in which human-computer interactions will be normal and

My general approach here is concrete experiential[3] in a way intended to go past the postmodern/modernist debate by entering into the wider implicit contexts of our interactions with computers. This approach also enlists adopt an antireductive approach to science and logic[4] that exhibits how their often overly simple models and merely formal precisions can also lead us to a richer understanding of the very factors they exclude. Accordingly I focus on the inherent losses of meaning that occur as the complex experiential richness of a person either goes right past the program's computationally limited capacity to process such a full range of meanings or simply fails to receive an adequate response from the computer, again because the model embedded in the program is inherently limited in its capacity to generate a fully engaging range of responses. Such losses of meaning occur even though the computer program provides extremely useful symbolic manipulations, thus eliciting such telling shifts in attitude as Kasparov's. Similar results that enter into crucial implicit contexts and thereby restore a wider, antireductive, intricacy will be expressed as specific claims about the experiential character of human-computer interactions and recollected in a list at the end of the essay. These results restore factors reductively eliminated or simply unexpressable in computational abstractions.

---

pervasive. In the long run, these newly transformed virtues might far outweigh the abusive effects, but only if we understand the limits and pitfalls of computationally generated symbols.

[3] This term is taken from Eugene T. Gendlin's work; see Gendlin, *Focusing-Oriented Psychotherapy* (New York: Guilford, 1996), and *Experiencing and the Creation of Meaning* (Evanston: Northwestern University Press, 1997); David Michael Levin, ed., *Language beyond Postmodernism*; (Evanston, Ill.: Northwestern University Press, 1997).

[4] Here following William C. Wimsatt; see Wimsatt, "Reductionistic Research Strategies and their Biases in the Units of Selection Controversy," in *Scientific Discovery, Volume II: Case Studies*, ed. T. Nickles. (Dordrecht: Reidel, pp. 213-259), "The Ontology of Complex Systems: Levels, Perspectives and Causal Thickets," *Canadian Journal of Philosophy* supplementary volume #20, ed. Robert Ware and Mohan Matthen, pp. 207-274, 1997), and "Emergence as Non-Aggregativity and the Biases of Reductionism(s)," in *Natural Contradictions: Perspectives on Ecology and Change* [Festschrift for Richard Levins], ed. J. Haila and P. Taylor, (SUNY Press, 199?).

My concern for the character of human-computer interactions gradually developed over the course of over two decades of working with people from many different walks of life as they faced several generations of new technologies in the workplace.[5] As I was employed in requirements definition, software development, and user support, I found I was really interested in the people involved: how they adapted to and took advantage of the technology, how they created meaning in interaction with it. The problematic in this essay will be to develop a theory of meaning loss in a hermeneutic context, as distinct from detailing a "human factors" approach. In contemporary practice, those losses are real and encountered daily by large numbers of people - and yet the experiential shifts of great importance to the people involved are not well addressed by a human factors concentration on the effectiveness and ergonomics of the user interface.

The basic humanistic interpretation of computer programs I adopt here is that they are 'symbolic' tools in the tradition of dictionaries, encyclopedias, concordances, etc., only computer programs work by adding computationally generated symbolic artifacts or merely "computed meanings" to whatever humanly generated meanings also happen to be recorded in such programs' digital medium and presented along with them.[6] Such "computed meanings" are strictly limited in their significance: they subsist

---

[5] I held various positions, from Information Management Specialist to Associate Director of Academic Computing, at the University of Chicago from 1975 to the present. I have also taught in the University's departments of Computer Science and General Studies in the Humanities in College since 1985.

[6] The shift in question is from the literate modes of human-human discourse practiced since the invention of writing to the artifactual or "informated" modes of human-computer-human discourse emergent with the invention of the artificial symbol systems of first-order logic and the artificial languages of computer programs. This shift is taking place not only for linguistic symbols, but for all modalities of symbolization including the visual, aural, and tactile as everything we capture in some medium becomes a potential symbol subject to discrete manipulation. Very quickly all our modes of communication are either becoming directly embedded in computerized symbol systems or in recurrent interaction with them. The term "informated" was developed as a broadly descriptive property of

as output generated by encapsulated semantic worlds or models encoded within the program. A computer running a program is really an object separable from its environment, an artifact or “black box”, containing algorithmic mechanisms for manipulating symbols usefully as it mediates between its input and output.

A great deal of work in philosophy, cognitive science, linguistics, and artificial intelligence confuses the basic fact that computers are external objects equipped with mechanisms, albeit very complex ones, for manipulating data and symbols, with the sometimes useful metaphor that "people are computers." Metaphors linking natural things and artifacts are a time-honored tradition, filled with both brilliant successes and dangerous failures. Plato compared the cosmos to a weaver's spindle; Descartes compared animals to automatons. Thinking of people as machines - or even worse of machines as more valuable than people - caused a great deal of human misery during the Industrial Revolution. We did not, however, on that account throw out machines; rather, we changed the main way we related people and machines. Accordingly, there is no need for us to get lost in the mistaken belief that people are "only computers" even as we use computer programs as tools to study ourselves and the world around us. A concern for the character of human-computer interactions necessarily leads us to develop interdisciplinary hermeneutics and semiotics of 'logical meaning creation' to mediate between fully human symbolization and engineered artifacts designed to manipulate symbols in useful ways.

Two observations about these human-computer interactions emerged quite clearly from my consulting work: 1) people naturally and unconsciously give their full

---

computerization by Shoshana Zuboff, *In the Age of the Smart Machine: The Future of Work and Power* (New York: Basic Books, 1988). For a global summary of modes of "learning, knowing and communicating," see Robert Beck, "The Future of Imaging Science." Preprint for *Advances in Visual Semiotics*, 1994.

meanings to the computer program as if it were another person, and 2) a definite artisanal knowledge for working with computers is developing that is not yet widely distributed in society and that is quite distinct from arts appropriate to other technologies such as industrial machines. In this paper I will deal primarily with the first and will only touch on the second only briefly, in its implications for the design of human-computer interactions with regard to the loss of meaning.

Let’s label my first observation the phenomenon of 'symbolic immediacy,” by which I mean the immediate re-cognition of already known meanings behind common words or images. In the way many readers would give immediate significance to "We hold these truths to be self-evident, that all men are created equal …" as a statement carrying deep and already known meanings, people give immediate recognition to computer-generated symbols. This attribution effect is well known to cognitive scientists: any number of studies demonstrate that even millisecond flashes of syllables will precondition which of several meanings a person will attribute to a full word. For our purposes, what is most salient about that fact is that computer-generated symbols can trick us into immediately attributing meaning as though they came from another person. This cognitive phenomenon is so strong that it resists further inquiry into its variations, particularly those of human-computer interactions. To that end I will employ the paired themes of anthropomorphism and anthropocentrism, two historically varied attitudes people have adapted toward the world of animals and things around us.

'Symbolic immediacy', however, is not a closed phenomenon. Just because we give immediate meaning to computer (or other human or artifactual) presented symbols in no way precludes prior, present, or further interactions with the meanings evoked. Symbolic immediacy in interaction with computers <u>is</u> situationally dependent and can be addressed in interdisciplinary, typical and specialized, and even cross-cultural ways by arts and sciences adapted to the interactive situation. This is not to say that human-computer interactions intrinsically tend toward such further interactions in ways that transform them for the better. If we want to encourage a transformation for the better, arts have to be explicitly formulated from relevant disciplines and why situationally generated artisanal knowledge must be proactively assimilated to such arts.

My long-term goal is to develop concepts and liberating arts that allow us to gain and regain a reflexive awareness of our symbols as separable from the meanings we attribute to them as a stage in the wider processes of experiencing meaning. I have previously formulated a system of arts for the context of computer programming consisting of requirements definition, algorithm construction, program composition, and program refinement.[7] In my consulting, I have adapted these arts to the specific resources and needs of the people involved.

My immediate project here is to make a few experiential observations and interpret them in light of the concepts of anthropomorphism and anthropocentrism via an extended analysis of Joseph Weizenbaum's ELIZA program in the context of actual psychotherapeutic transcripts. On the basis of this analysis, I develop a typology of symbolic demeaning – or, in other words, the ways meaning is lost in human-computer interactions - along with some experiential claims. Included in that analysis, I give a brief indication of some implications of this typology as goals for ameliorative software design. I conclude with a résumé of experiential and antireductive points and claims made throughout the essay. The overall gist of my argument is not so much to focus on the successful application of relevant arts to human-computer interactions as to examine what happens when such arts are not properly at work.

Anthropomorphism and ELIZA

In January, 1966 Joseph Weizenbaum published a masterful article on computational linguistics entitled "ELIZA--A Computer Program for the Study of Natural Language Communication between Man and Machine," in the leading journal of the computing community, *The Communications of the ACM (Association for Computing Machinery)*.[8] Weizenbaum's article epitomized the creativity, the power, and the dependence on anthropomorphism of the Strong AI program.

Supported in part by Department of Defense (DOD) grant moneys to MIT, ELIZA was one of the earliest programs using the MAC time-sharing system that permitted a real time "natural language conversation between man and computer" to be the user interface for ELIZA. The new time-sharing system allowed a free exchange of symbols between user and "computer screen," thereby satisfying one of the descriptive features of Turing's test with much greater spectacle. When the program quickly responded with symbols on the screen, it felt as if someone was typing messages to you. With the intense attributions of intelligence surrounding computers as “electronic brains” at that time, it was inevitable that people would find the computer's output meaningful to themselves. ELIZA was perceived as personally responsive.

Another significant innovation was the use of "editable scripts" consisting of keywords and transformation rules[9] as the primary data structure for ELIZA's linguistic

---

[7] William H. Sterner, "Computer Programming, Curricular Pluralisms, and the Liberal Arts," presented at "Systematic Pluralism: An Interdisciplinary Conference," University of Nebraska-Lincoln, April, 1990. Copies available on request.

[8] See Appendix 1 for a copy of the article. Joseph Weizenbaum, "ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine," *The Communications of the ACM (Association for Computing Machinery)* Vol. 9:1, (January 1966): 36-45.

[9] These are not the same as Chomsky’s generative grammars, but rather much simpler, entirely mechanical word manipulations. See the copy of Weizenbaum’s paper in an appendix.

interactions. Such scripts contained linguistic data separate from the program's conversational engine and, hence, could be modified or extended without having to change the engine part of the program. At the time the conversations that could be achieved with this data structure seemed to be unlimited – or, at least, possible limitations were not well understood. It gave all the appearance of being open to whatever semantic content and patterns of grammatical transformation might be desired. The potential for an indefinitely large number of possible inputs was not taken to be a problem. Getting the right responses to a given input was seen as "a mere matter of programming" the right script.

The operation of the program begins with the analysis of "input sentences" by means of "decomposition templates" triggered by scripted keywords found in specific locations in the text. For example, the input "It seems that you hate me," with keywords "you" and "me," would be broken down according to segments of keywords and segments without keywords:

"(1) It seems that (2) <u>you</u> (3) hate (4) <u>me</u>".

This input then could correspond to a decomposition template of "(0 YOU 0 ME)" where "0" stands for an indefinite number of words. Having gotten this far in "understanding" the input, the program must associate the matching template and keywords with a "reassembly rule" that can be used as a template for ELIZA's response. The specific rule "(WHAT MAKES YOU THINK I 3 YOU)" where the "3" identifies that the third component of the input, i.e., "hate," then yields the output, " WHAT MAKES YOU THINK I HATE YOU," From the user's point of view the exchange cycle takes on an somewhat natural aspect:

User: It seems that you hate me.

ELIZA: WHAT MAKES YOU THINK I HATE YOU[10]

From the programmer's point of view all this was quite plausible as "natural language conversation" for the two reasons: (1) because the syntactically clear decomposition and reassembly rules, along with the semantically discrete keywords, were taken to be an adequate filter for natural language, and (2) because the goal in mind was the construction of encapsulated domains of conversational content. Indeed, there are several additional features of the program's operation that enhance this exchange cycle. For instance, many varied decomposition templates could be constructed. More than one reassembly rule could then be associated with each decomposition, and various keyword rankings, memory tricks, statistical patterns, and other secondary mechanisms could be invoked to choose among different reassembly rules. Technologically, this machinery provides for a fascinating array of linguistic manipulations that seem endlessly suggestive. The further tinkerings caught sight of in this vista, however, are really anthropomorphic projections of the nuanced meanings experienced by the programmer and immediately given over to words stored in the program.[11]

Weizenbaum reports the following as a "typical conversation" with ELIZA:

User: Men are all alike.

ELIZA: IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

---

[10] Weizenbaum, “ELIZA,” pp. 37-39.

[11] I use “anthropomorphic” here in J. S. Kennedy's more rigorous sense of the intrinsic embeddedness of metaphors in all natural language use including scientific description. Kennedy, J. S., *The New Anthropomorphism* (Cambridge: Cambridge University Press, 1992).

Well, my boyfriend made me come here.

YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.

I AM SORRY TO HEAR YOU ARE DEPRESSED

It's true. I am unhappy.

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

I need some help, that much seems certain.

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

Perhaps I could learn to get along with my mother.

TELL ME MORE ABOUT YOUR FAMILY

My mother takes care of me.

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.

YOUR FATHER

You are like my father in some ways.

WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't want me to notice that.

WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE

You don't argue with me.

WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

My father is afraid of everybody.

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

Bullies.

> DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE[12]

---

[12] Weizenbaum, “ELIZA,” pp. 36-37.

We will return to this script shortly, but first we must grapple with Weizenbaum's exposition of ELIZA as an artificial intelligence. Masterful as his AI and heuristic programming advances were, Weizenbaum's rhetorical placement of ELIZA within a client-centered therapy framework was brilliant. Quite sensitive to the role of "magic" with in machines "made to behave in wondrous ways, often sufficient to dazzle even the most experienced observer,"[13] Weizenbaum found in client-centered therapy an almost perfect framework to take best advantage of our inevitable anthropomorphism in linguistic interaction and thereby conceal the program's conversational limitations. As he purposefully admits with the intention of making the magic crumble:

> ELIZA performs best when its human correspondent is initially instructed to "talk" to it, via the typewriter of course, just as one would to a psychiatrist. This mode of conversation was chosen because the psychiatric interview is one of the few examples of categorized dyadic natural language communication in which one of the participating pair is free to assume the pose of knowing almost nothing of the real world. If, for example, one were to tell a psychiatrist "I went for a long boat ride" and he responded "Tell me about boats", one would not assume that he knew nothing about boats, but that he had some purpose in so directing the subsequent conversation. It is important to note that this assumption is one made by the speaker. … The speaker further defends his impression (which even in real life may be illusory) by attributing to his conversational partner all sorts of background knowledge, insights and reasoning ability.[14]

And what was it that motivated this Rogerian guise?

> From the purely technical programming point of view then, the psychiatric interview form of an ELIZA script has the advantage that it eliminates the need of storing *explicit* information about the real world.

In other words, the initial purpose was to lessen the burden on his role of playing Henry Higgins to Eliza Dolittle. But that got out of hand as Weizenbaum focused in on how all AI depends upon our unwitting anthropomorphic generosity.

---

[13] *Ibid*., p. 36.
[14] *Ibid*., p. 42.

ELIZA was immensely successful. In 1966 Weizenbaum already reported that "some subjects have been very hard to convince that ELIZA (with its present script) is *not* human." No less an AI luminary than Herbert Simon expressed the opinion in 1966 that

> If the [ELIZA] method proves beneficial, then it would provide a therapeutic tool which can be made widely available to mental hospitals and psychiatric centers suffering a shortage of therapists … several hundred patients an hour could be handled by a computer system.[15]

Psychiatrist Mark Colby among others got on the bandwagon. Interest became so heated that the press and even congressmen picked up on the fantasy that ELIZA could substitute for human therapists; government funding was sought to finance widespread use of the ELIZA program to cut the costs of psychotherapy. We can be grateful that Weizenbaum himself was motivated by the runaway response at the time to write about the dangers of overestimating "computer power" as "human reason." His book *Computer Power and Human Reason: from Judgment to Calculation* was published in 1976.

And so, much to his credit as he disassociated himself from the "artificial intelligentsia," Weizenbaum wound up heading his own earlier warning expressed in his ELIZA article:

> The whole issue of the credibility (to humans) of machine output demands investigation. Important decisions increasingly tend to be made in response to computer output. The ultimately responsible human interpreter of "What the machine says" is, not unlike the correspondent with ELIZA, constantly faced with the need to make credibility judgments. ELIZA shows, if nothing else, how easy it is to create and maintain the illusion of understanding, hence perhaps of judgment deserving of credibility. A certain danger lurks there.[16]

---

[15] Herbert Simon as quoted in Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (San Francisco: W. H. Freeman and Company, 1976), p. 180.

[16] Weizenbaum, "ELIZA," pp. 42-43.

Weizenbaum had discovered from within the AI project exactly the premise behind all the attempts to create artifacts that successfully imitate intelligence, namely, that we find meaning and respond to symbols that look as though humans had created them. And even if the symbols do not look quite right, we tend to forgive and stretch to maintain our belief in machine intelligence.

At this point all sorts of controversy rushes into the discussion. Hubert Dreyfus acknowledges that Weizenbaum was the first, and among the few of the Strong AI camp, to recognize these issues and actually change his stance in response. Yet Dreyfus takes even Weizenbaum to task for not going far enough to find the ultimate dividing lines between everyday understanding and natural language communication, on the one hand, and a context-free[17] logicality and formalizable intelligence, on the other.[18] Many Good Old-Fashioned AI (GOFAI) advocates such as Marvin Minsky, Roger Schank willfully carry on as if none of this applies to their projects. They have work to do that has never been done before and for them that is warrant enough. And many new cognitive scientists, philosophers, linguists, biologists, chaos theorists, computer scientists, physicists, and so forth. join the fray from all sorts of angles. I contend that much of this controversy is fruitless simply because it is entrenched in the backwaters of the "mind as computer" metaphor whenever it entails the reduction of human meaning creation to some computational mechanism. Instead of entering this confusion further, I will henceforth focus on characteristics of human-computer interaction in which computers are treated as mechanized symbolic artifacts or "artifactual symbol systems."

---

[17] See footnote 111, p. 98 below for a further argument about the significance of this term.

[18] Hubert L. Dreyfus, *What Computers **Still** Can't Do: A Critique of Artificial Reason* (Cambridge, Mass: MIT Press, 1992), pp. 65ff., 219ff., and 314.

My agenda from here on in is first to explore how ELIZA breaks down, and then to examine how interactional complexity looks in a real therapeutic conversation. With that contrast in mind, I propose some categories of *symbolic demeaning*, or dialogic failure to which human-computer interactions are subject; then I explore the implications for program design that might in part respond to those categories; and, finally, I return to a twentieth-century emblem of the entire problematic. The My aim throughout is not to avoid controversy per se, but to address concrete issues that get lost in it.

ELIZA and Natural Language

Consider the following three examples of conversation while keeping the ELIZA exchange in mind. The first is an excerpt from a hypothetical exchange between a judge in a Turing test and a computer program named Chinese Room imagined by the philosopher Daniel Dennett.[19] The second is from a transcript of a training session between Alex the gray parrot and two trainers reported by Irene Pepperberg. And the third is an anecdote about Panbanisha, the bonobo ape, reported by Jessica Heineman-Pieper from a talk by Sue Savage-Rumbaugh. Together the four exchanges give an interesting range of human to non-human conversations.

Dennett

> Judge (of Turing test): Did you hear about the Irishman who found a magic lamp? When he rubbed it a genie appeared and granted him three wishes. "I'll have a pint of Guinness!" the Irishman replied, and immediately it appeared. The Irishman eagerly set to sipping and then gulping, but the level of Guinness in the glass was always magically restored. After a while the genie became impatient. "Well, what about your second wish?" he asked. Replied the Irishman between gulps, "Oh well, I guess I'll have another one of these!"

---

[19] A Turing test is some situation in which a computer is anonymously grouped among people asked to engage in exchanges with a judge who's task is to identify the computer. If the judge fails to separate out the computer from the people, the computer passes the Turing test as in some fashion equivalent to a human respondent. The procedure was first proposed by Alan Turing in 1950.

> Chinese Room (computer program): Very funny. No, I hadn't heard it -- but you know, I find ethnic jokes in bad taste. I laughed in spite of myself, but really, I think you should find other topics for us to discuss.
>
> J: Fair enough, but I told you the joke because I want you to explain it to me.
>
> CR: Boring! You should never explain jokes.
>
> J: Nevertheless, this is my test question. Can you explain to me how and why the joke "works"?
>
> CR: If you insist. You see, it depends on the assumption that the magically refilling glass will go on refilling forever, so the Irishman has all the stout he can ever drink. So he hardly has a reason for wanting a duplicate, but he is so stupid (that's the part I object to) or so besotted by the alcohol, that he doesn't recognize this, and so, unthinkingly endorsing his delight with his first wish come true, he asks for seconds. These background assumptions aren't true, of course, but just part of the ambient lore of joke-telling, in which we suspend our disbelief in magic, and so forth. By the way, we could imagine a somewhat labored continuation in which the Irishman turned out to be "right" in his second wish after all--perhaps he's planning to throw a big party, and one glass won't refill fast enough to satisfy all his thirsty guests (and it's no use saving it up in advance--we all know how stale stout loses its taste). We tend not to think of such complications, which is part of the explanation of why jokes work. Is that enough?[20]

Alex the Gray Parrot

Features of human-animal relations that any ethical analysis will need to take into account are precisely those of human-animal communication. What, in fact, do the linguistic bridges we have built with primates and even parrots allow us to communicate? The following exchange between Alex the gray parrot and two trainers, "I" and "B," given by Pepperberg is instructive:

Excerpt of a Model/Rival (M/R) Training Session

I (acting as trainer) Bruce, what's this?

B (acting as model/rival) *Five* wood.

---

[20] Daniel C. Dennett, *Consciousness Explained* (Boston: Little, Brown and Company, 1991), pp. 436-437.

I That's right, *five* wood. Here you are … *five* wood. (Hands over five wooden popsicle sticks. B begins to break one apart, much as Alex would.)

A 'ii wood.

B (now acting as trainer, quickly replaces broken stick and presents the five sticks to Alex) Better … (briefly turns away then repositions himself in visual contact with Alex) … How many?

A No!

B (Turns from Alex to establish visual contact with the PI.) Irene, what's this? (presents sticks).

I (Now acting as model/rival) 'ii wood.

B Better … (turns then resumes eye contact) … How many?

I *Five* wood (takes wooden sticks) … *five* wood. (Now acts as trainer, directs gaze to Alex, and presents sticks to him) … How many wood?

A Fife wood.

I OK, Alex, close enough … *fivvvvve* wood … Here's *five* wood. (Places one stick in the bird's beak and the other within his reach.)

Note. The aim of the session was to review and improve pronunciation of the label "five."[21]

Despite the impressive range of concept discriminations Alex appears to achieve (such as "What's here?", "What color?", "What shape?" and the more difficult "What's same?", "What's different?"), certainly the full expressiveness of human natural language is not at work in conversations with Alex. This is some sort of simpler grammar, or "artificial language" tool for establishing communication, perhaps even made possible by the developments in generative grammar and artificial language construction with their ranges of different expressive capacities. But the issue of animal

---

[21] I. M. Pepperberg, "An Interactive Modeling Technique for Acquisition of Communication Skills: Separation of a 'Labeling' and 'Requesting' in a PsiHacine Subject," *Applied Psycholinguistics* 9, (1988): 163.

consciousness does not permit us to encapsulate the communicative interactions within a "pure research" context just because of their evident simplicity.

Panbanisha the bonobo chimpanzee

An exemplary anecdote reported by Jessica Heineman-Pieper about one of Sue Savage-Rumbaugh's primates puts a definite edge to the point.

> In a standard "false belief test" experimental situation, a favorite object or food is placed in one of two containers by one experimenter who then leaves the room, and is then moved to a second container by another experimenter all within sight of a human child three to four years old. When asked where the first experimenter would think the object is, a four year old child would typically respond with the first container, indicating that the child has some representation of the first experimenter's beliefs. This is different for a three-year old human who will select the second container. The reported anecdote takes place in a more naturalistic setting. While on a walk in the woods with Panbanisha, one experimenter mentioned a desire for something, perhaps an M&M candy. The second experimenter said she had one right here and put it into a box with Pambanisha watching. Then the first experimenter went away for a little while. Again with the primate watching the second experimenter switched the container, and then asked where Panbanisha thought the first experimenter would look. The primate signed the first container. Then she further signed the question as to why the second experimenter was being mean to the first.[22]

With that autonomous question, made possible by years of training and human-primate interaction, the issue of trans-species ethics takes on a complexly different framework.

Is ELIZA just a technically unsophisticated precedent to Dennett's CR (Chinese Room) program? Is it (she) a better conversationalist than Kanzi's sister who wondered why the experimenter was being mean? Is the way that Alex makes

---

[22] From a talk by Sue Savage-Rumbaugh given at the "Toward a Science of Consciousness Conference," Tucson, Arizona, April, 1996, as reported in W. Wimsatt's Philosophy of Biology course that April by Jessica Heineman-Pieper.

discriminations just as mechanical as ELIZA's word manipulations? Or is there some overwhelming gap between the symbols produced by ELIZA and either CR or Kanzi's sister, not to mention a human therapist?

For all his forthrightness, Weizenbaum is not a solid guide here: a "typical" conversation with ELIZA shows a lot more brittleness. What follows are some results when statements taken from a real therapy session[23] are put into the original version of the program. That is, in the transcripts below I am submitting the client's side of an recorded session between a human therapist and client to the ELIZA program and then continuing the exchange with the next client statement <u>regardless</u> of ELIZA's response. We will see later that the dynamics of a genuine therapy session work to keep ELIZA going further successfully than if the client were to directly respond to ELIZA as a therapist. So, at first, let us assume that the client in the form of her transcribed utterances is especially willing to suspend disbelief in the computer therapist's fallibility and just keeps pushing on with her interactions. My own comments are given within [square brackets]. (This newer version of ELIZA stopped shouting at its clients in all upper case letters, but is otherwise identical with Weizenbaum's first published version.)

1) ELIZA: How do you do. Please tell me your problem.
Client: I was thinking about -- on my way over -- I don't seem to think a hell of a lot of myself.
E: Your way over you don't seem to think a hell of a lot of yourself?
C: Well, uh ...
E: Lets discuss further why your way over you don't seem to think a hell of a lot of yourself.

[Most people would think this therapist has something of a speech problem. Maybe she drops words so that "Your way over you" was really intended to be "On

---

[23] Gendlin, Eugene T., *Focusing-Oriented Psychotherapy*, (New York: Guilford, 1996), pp. 28-32.

your way over you were thinking …" But she invites further discussion even if she is going "way over" on you. Let's keep going.]

2) C: I had a dream -- I was alone with this guy, ah -- and the dream was real nice, it was a real nice relationship.
E: What does that dream suggest to you?
C: When I thought about it next day I thought, why don't I have a real one!
E: Why do you ask?
C: I don't think he could really see anything wrong with me.
E: Don't you really think he could really see anything wrong with you?

[Yet again, maybe the therapist isn't entirely sympathetic. Doesn't ELIZA understand how difficult it is for the client to get a real relationship going? Even worse, the therapist seems convinced that the guy "could really see anything wrong" with the client. Isn't that the client's worst fear, that the guy could see right through her to all her faults? But maybe this isn't a Rogerian "unconditional positive regard" kind of therapy after all. Perhaps it's some sort of "harsh reality" approach.]

3) C: I was also thinking why I was absent in school so much.
E: You say you was also thinking why you was absent in school so much?
C: When it comes to the end of the line I don't have a paper, I hold back.
E: Does that question interest you?
C: I get jittery and then I pull away from it.
E: Can you elaborate on that?
C: Yeah. I have all these excuses about why I never do my best, uh ...
E: Does that suggest anything else which belongs to you?

[Well there's that odd speech pattern again -- "You say you was also thinking why you was absent." Assuming the client is sincere, then absence and not turning in papers seems a real problem for her. Not much support to be found in: "Does that question interest you?", which carries more the sense that the therapist couldn't be more bored. Then the therapist stalls for time by asking for further elaboration and hints that the client always has excuses. Armed with this new information we get knocked for a loop! Our compliant client might think, "Well, I guess my excuses belong to me. What does that remind me of? What was I talking about?" Carrying on, we get:]

4) C: Yea, yea. Uh -- I pull back.
E: Do you say uh you pull back for some special reason?
C: The jittery is more a surface than the pull back.
E: Earlier you said your best.

[Things seem to be breaking down on both sides here. ELIZA has invoked a memory mechanism bringing back a previously stored input word "best," most likely because it failed to find any keywords in the client's text.]

5) C: The jittery comes when part of me says well, you know, you really have to do it now.
E: We were discussing you - not me.

[That's a bit snappy on E's part. Perhaps "the therapist in the machine" is getting a bit annoyed with the client with all those embedded "you's".]

6) C: No, the jittery is a result.
E: Are you saying no just to be negative?
C: Well I think it's -- ah -- that I don't want to test myself.
E: Why don't you want to test yourself?
C: And I'm afraid, ah, the bad things will be confirmed.
E: Is it because you are afraid that you came to me?

[We have definitely shifted to a more confrontational mode. And, apparently, the therapist cannot imagine any other reason why the client came. It's not clear just what the client is getting out of the interaction by this point.]

The surprising thing is how well ELIZA did with a real transcript (i.e., one that kept submitting the sequence engendered by a real therapist, no matter what ELIZA replied). Things run aground much more quickly if the person actually pays attention to what ELIZA has to say. Going back to some of the rough spots and this time with me (S) “intervening” and responding directly to ELIZA when things bog down, we get a less satisfying dialog.

1b) ELIZA: How do you do. Please tell my your problem.
Client: I was thinking about - on my way over - I don't seem to think a hell of a lot of myself.
E: Your way over you don't seem to think a hell of a lot of yourself?
S: I'm not way over myself, I was just not thinking a hell of a lot of myself.
E: Is it because you are not way over yourself that you came to me?
S: It really isn't that bad - sure I have low self esteem, but I'm not way over myself.
E: You say it really isn't that bad sure you have low self esteem?
S: Yeah, I don't seem to think a hell of a lot of myself.
E: Don't you really seem to think a hell of a lot of yourself?
S: No I don't. You are making that up.
E: Are you saying no just to be negative?
S: No, it's because you are insulting me.
E: You are being a bit negative.
S: No I'm not.

E: Why not?

[Actually, by this point I was feeling pretty negative.]

3) C: I was also thinking why I was absent in school so much.
E: You say you was also thinking why you was absent in school so much?
S: It sounds like you could use some extra schooling as well.
E: We were discussing you - not me.
C: When it comes to the end of the line I don't have a paper, I hold back.
E: Why do you ask?
S: Huh, I didn't ask anything.
E: Can you elaborate on that?
S: When I talked about not having a paper you thought that was a question.
E: Does that question interest you?
S: What question?
E: What is it you really want to know?

[Whatever is going on here is beyond me as a putative client.]

An easy temptation is to get into explicitly trying to trip up ELIZA's decomposition rules, but we will skip that exercise.[24] While the script does have some capacity to encode conversational gambits and the occasional therapeutic turn of phrase that throws the client back on her own projections, ELIZA really has an extremely small knowledge base. That limitation of course was used to set the horizon for the next generation of AI efforts -- such as Expert Systems in the 1980s and recently Doug Lenat's Cyc database of commonsense knowledge.

But before we deposit that promissory note, we should pause to remember that, historically, the original published conversation was the sort of discourse that Herbert Simon immediately thought might "provide a therapeutic tool which can be made widely available to mental hospitals and psychiatric centers suffering a shortage of

---

[24] The reader is encouraged to try ELIZA and KantPro for themselves. A version of ELIZA is available on line at http:/www-ai.ijs.si/eliza/eliza.html, and I am happy to provide a copy of Kantpro to people on campus. It is a Macintosh program.

therapists."[25] It is also instructive to compare it to the creative heights of fictional conversation achieved by the philosophical luminary Daniel Dennett on behalf of his promised CR program.[26] Again, heavy anthropomorphism is at work in disciplinary discourse, -- in computer science, in one case, and in philosophy in the other.

The key point for us is that although it is obvious just how trivial and impoverished computer generated speech is while we are looking at the symbol generating mechanisms of ELIZA's decomposition and reassembly rules, we immediately forget those limitations when presented with English words on the screen. Precisely as computerized re-enactments of conversation incrementally improve, we are in increasing danger of forgetting the canned, imitative mechanisms that drive them and that do so in ways not at all "natural" after the fashion of natural language.

Descent into Human Complexity

We could turn to the discursive heights of literature, science, or philosophy for comparisons with ELIZA's computer-generated symbolizations, but to do this properly we would need to have an architectonic poetics at hand that embraced the range of symbolic artifacts as works of science and art. Steps in that direction have been taken, notably by Herbert Simon in a different guise[27] and by Nelson Goodman in his *Languages of Art*. But major hurdles remain, with the poetics of "interactive fiction" and

---

[25] Just what Weizenbaum and, apparently, Simon thought was "typical" of a psychotherapeutic conversation is itself an interesting question. The published mock interaction exhibits a good deal of what psychotherapy was thought to be in the 1960s. It is clear that the interactively “thick” context of therapy was not well understood in popular culture.

[26] To carry the "replacement of humans" theme a step further, see a computerized version of Kant's philosophical discourse in Appendix 2. On the other hand, for a human appropriation of ELIZA's computerized discourse, see the excerpt from David Lodge's *Small World* (New York: Macmillan Publishing Company, 198), pp. 242-247 and 308-310.

[27] Simon's *Sciences of the Artificial* can certainly be characterized in this manner.

its innovations on an author's dramatic manner of imitation (in the Aristotelian sense) as just one of them.[28]

Fortunately we can find a narrower path more in line with the ELIZA theme. Let us turn, then, to the more interactively structured and relatively discrete exchanges of an actual human client and therapist. Perhaps with the previous exercise in mind, we will be able to notice a few features of an intensely subjective symbolic interaction with a greater clarity than Simon or Dennett is committed to attaining.

When we move into a genuine psychotherapeutic situation, there are sensitive issues of responsibility that demand our serious attention. With Darwin's nineteenth-century notion of moral sense, alternative courses of action are evaluated by

> [an] inward monitor [that] would tell the animal that it would have been better to have followed the one impulse rather than the other. The one course ought to have been followed: the one would have been right and the other wrong;[29]

Thus for "normal" individuals (with sufficiently advanced intelligence), we would expect that they

> would have an inward sense of possessing certain stronger or more enduring instincts, and others less strong or enduring; so that there would often be a struggle which impulse should be followed; and satisfaction or dissatisfaction would be felt, as past impressions were compared during their incessant passage through the mind.[30]

This mechanism consequently allows for society to find moral criteria for assigning praise and blame, punishment and reward, depending upon an individual's ability to behave according to community standards of right conduct. However, as the contemporary of Darwin, philosopher John Stuart Mill elucidates, this series of

---

[28] The recent work of Selmer Bringsjord and David A. Ferrucci *Artificial Intelligence and Literary Creativity: Inside the Mind of BRUTUS, a Storytelling Machine* (Mahwah, New Jersey: lawrence Earlbaum Associates, 2000) looks to be a fascinating discussion of how the authors have tried hard to construct a program that tells stories and failed.

[29] Charles Darwin, *The Descent of Man, and Selection in Relation to Sex* (Princeton: Princeton University Press, (1872) 1981, p. 73-74.

judgments by community and individual leads us to a situation in which anyone acting or behaving from a wrong-minded disposition, of consequence

> places his mind out of sympathy with the rest of his fellow-creatures, and if they are aware of his disposition, becomes a natural object of their active dislike. He not only forfeits the pleasure of their good will, but the benefit of their good offices, except when compassion for the human being is stronger than distaste toward the wrong-doer;[31]

Setting aside the overlapping cases, psychologically disturbed people fall into this latter category – that is, they deserve our compassion -- rather than into the overtly moral categories of weak character or criminal disposition. Yet even such a cleaned-up division by no means removes the general hostility directed toward noncriminal neurotic, psychotic, or otherwise disturbed individuals as people putatively not living up to their duties to maintain and enhance society.

This twentieth-century socially constructed "role" of the psychotherapist still carries with it the historical decisions of our society to make a specialized "caring relationship" part of the intended structure of society. It is considered institutionally and socially legitimate to help people undergoing difficulties with other people and with themselves while also continuing to make moral judgments on people in therapy. The place of social authority within psychotherapy is controversial today. Our expert source, Eugene Gendlin, is very careful to not confuse the legitimacies and responsibilities of a professional caring relationship with the actual interactions between client and therapist. He distinguishes the "administrative role" of the therapist with its unequal power relation between client and therapist from the fundamental question of "Did a

---

[30] *Ibid*., p. 73

[31] John Stuart Mill, *an Examination of Sir William Hamilton's Philosophy*, Ch. XXVI (New York, 1884), vol. II, pp. 228-289. Quoted in Richard McKeon, "The Development and the Significance of the Concept of Responsibility," *Revue international de philsophie* 34 (1957, fasc. 1): p. 21. Reprinted in Richard Mckeon, *Freedom and History and Other Essays* Zahava, K. McKeon, ed. (Chicago: University of Chicago Press, 1990), p. 78.

therapeutic process happen for the client?" Methodically imposing "socially acceptable" ideas and actions on the client is the antithesis of a therapeutic process. Instead the client must be free to express themselves honestly in the "real life-relationship between two people," where "All feelings are welcomed, but possible actions are highly restricted." Society has legitimated the caring, but only in a generic way. In deep contrast, Gendlin shows that individuals bring their own embodied self-organizing order to their experiencing; it is not all social forms of behavior imposed on biological mechanisms.[32] The interactively emergent specifics of individuals must be taken as the basis for the therapeutic relationship whatever the wider social goods. The lack of a widespread societal understanding of these infinities of nuance in human complexity make it easy for us to mistake computer generated symbols for what people actually mean and experience through their embodied symbolizations.

It is precisely because of these requirements for implicit as well as strictly conceptual human-human communication are so exacting, that the therapeutic interaction offers an exemplary case to compare with human-computer interactions. At the one extreme, the possibility of a “strictly logical computer program” that nevertheless could provide empathic responses was immediately attractive as a more "scientific" and "objective" solution to the social and individual problems of the day. The immediate enthusiasm in the 1960s and 1970s at the prospect of an ELIZA-type program provides evidence of that attractiveness.[33] ELIZA was frequently accorded the status of a "socially objective" authority simply because it was constructed from

---

[32] Gendlin , *Focusing-Oriented Psychotherapy*, p. 303. See also “The Small Steps of the Therapy process: How They Come and How to Help Them Come,” in Lietaer, G., Rombauts, J., Van Balen, R. (1990) *Client-centered and Experiential Psychotherapy in the Nineties*. Leuven, Gelgium: Leuven University Press. (Especially the section “Human Nature: Imposed Form Versus and Order of Steps,” pp. 208-216.)

[33] Weizenbaum, *Computer Power and Human Reason*, ...

logical rules. At the opposite pole, provided by the twentieth-century professions of psychotherapy, we have perhaps the most extremely proper or legitimate use of anthropomorphic language, namely, in opening up of a deeper, healthier humanity through the interactive process with a disturbed client who has somehow become alienated from the full functioning of their individual human capacities. The crux of the difference between these two extremes is in the extent to which we can struggle to realize a more objectively fulfilling humanity through the disciplined and artful use of language. Rather than turning to objectivity found in the logic of a program, this use avoids a danger of unscientific subjectivity through its own pragmatic discipline – a discipline arising from the intrinsically anthropomorphic properties of natural language by so enriching the client's world of meanings as to foster psychological growth. This was the task many AI and other professionals implicitly assumed ELIZA's logic was up to. That is, if they even had an understanding of a genuine therapeutic relationship.

Carl Rogers's client-centered therapy arose as one of several American alternatives to Freudian psychoanalysis around the midpoint of the century -- Along with Harry Stack Sullivan's interpersonal theory of psychiatry, Rollo May's existential psychotherapy, Abraham Maslow's psychology of being, Fritz Perls' gestalt therapy, among several others. Along with these clinicians, Rogers sought an alternative theory of personality and therapeutic discipline. His was an emphatically democratic and pragmatically scientific approach. For Rogers, the overbearing authority of the psychoanalyst gave way to and became embodied in a more open and receptive therapist; patients, in turn, were viewed as clients seeking counseling rather than invalids requiring a medical diagnosis and treatment. Many issues pertaining to objectivity and scientific verifiability in the air at midcentury now presented opportunities for direct psychological research. This approach was more than mere scientific rhetoric. Rogers built a thriving intellectual community at the University of Chicago in the 1950s that coalesced around the University of Chicago Counseling Center. Here several dozen thinkers and psychologists worked under experimental regimens to determine therapeutic techniques that could be identified as fostering successful personality change.[34] Many concrete techniques for resolving the difficulties with constructing scientific theories of therapy were first developed here in aggressive empirical research programs. The finding that success in therapy could not be correlated with any one personality theory was especially noteworthy. The standard was raised even higher when this finding in turn encouraged a search for "process models" that could identify and study second-order interactional variables that emerged regardless of the personality theory used. Moreover, an integral part of the school's standard operating procedures was to include cross-therapist ratings of therapy events, which were made possible for the first time by tape recordings of therapy sessions. In short, Karl Popper's objections to Freud's supposed irrefutability were becoming assimilated as a structural part of the discipline even as the differences between human science and physical science were still being debated and reorganized.

The transcript that follows, one which I used as a source of input above, comes from a therapy session with Eugene Gendlin, who is considered to be Rogers's most theoretically adept student and who now heads a successful international school of psychotherapy. Gendlin received a joint Ph.D. in philosophy and psychology at the

---

[34] See over ten years of *Chicago University Counseling Center Discussion Papers* ed. John M. Butler (Regenstein Library, University of Chicago, call number: Bf 637 C6C53).

University of Chicago in 1958. This excerpt comes from Gendlin's recently published book, *Focusing-Oriented Psychotherapy*.[35]

---

[35] Gendlin, *Focusing-Oriented Psychotherapy* (1996), chapters 4 and 5, pp. 28-32.

	Transcript		Gendlin's first commentary
C1:	I was thinking about … on my way over … I don't seem to think a hell of a lot of myself.	C1	It's the beginning of the hour, and this is what she thought about on the way here.
T1:	So … you're asking … why do you have such a low estimate of yourself.		
C2:	Well, uh --	C2	What he said doesn't fit.
T2:	Or, maybe not asking, exactly.		
C3:	I had a dream … I was alone with this guy, ah (*silence*) … and the dream was real nice, it was a real nice relationship. When I thought about it the next day I thought, why don't I have a real one! I don't think he could really see anything wrong with me. I was also thinking why I was absent in school so much. When it comes to the end of the line I don't have a paper, I hold back. I get jittery and then I pull away from it.	C3	By "end of the line," she means when it's time to hand in the paper, or actually get involved with a man.
T3:	You're saying there is something similar about those two things.		
C4	Yeah. I have all these excuses about why I never do my best, uh …	C4	"Never do my best" -- that is, it will not be a real test.
T4:	You come right up to the line and then something holds back.		
C5:	Yeah.		
T5:	And "jittery" is the best word for it.		
C6:	Yeah, yeah. Uh … I pull back.	C6	She means that "jittery" is not the best word for it. "I pull back" is.

	Transcript		Gendlin's first commentary
T6:	Pull back is it.		
C7:	The jittery is more a surface than the pull back. The jittery comes when part of me says, "well, you know, you really have to do it now."	C7	She gets jittery when she thinks she will force herself to do it. But feeling jittery is not what prevents her from doing it.
T7:	So we don't know what pulls back, it's not the jittery that pulls.		
C8:	No, the jittery is a result.		
T8:	So we don't really know what the pull back feels like, what it is that wants to pull back.	C9	She is not willing to let go of what she is thinking and to sense what the pulling back feels like, as he invites her to do. Rather she repeats what she said in C3 and C4, which he does not want to hear, and did not respond to. She thinks she avoids a real test for fear that she will find out that she is not that brilliant, or that she is not attractive to men.
C9:	Well I think it's … ah … that I don't want to test myself. And I'm afraid, ah, the bad things will be confirmed.		
T9:	Can you feel the pull back, if you imagine yourself going ahead?		
C10:	Yeah, I can feel the pull back now. … The pull back is into weed, that's what it does.	C10	She can feel her desire, right now, to pull back into marijuana. She might mean she wishes she could smoke some right now. Or, she might mean that her pulling back often pulls her into dope.
T10:	Into weed.		
C11:	Marijuana, that's the perfect place to pull back.		
T11:	That's a perfect place to pull back to.		
C12:	Yea. But if I don't go to the line then I don't have to pull back.		
T12:	As long as you don't really go across the line, there is no testing of it, there is no proof, good, bad, and you're suspecting that you're afraid of actually finding out.		

Transcript	Gendlin's first commentary
C13: Right!	C13 This was a strong "right!" Finally the therapist attended to her view of the problem. Until now he ignored what she has been saying (because it was similar to what she had said before, which led nowhere.) But he should have responded as in T12 sooner.
T13: I was interested also in just the feel quality of it, for a minute you could feel the pull back.	
C14: Yeah, I could feel it.	C14 As she says this, more quietly, she seems to be sensing the pulling back right now.
T14: Let's just tap it lightly, and see what it turns up. (*There is a short silence.*)	
C15: Scared … it's like the world is going to bite me or something. (*Laughs*)	C15 This describes *the quality* of the unclear sense of the whole of it. It's "scared," and more exactly, this kind of scared.
T15: Um Hum. Yeah, yeah. (*More silence*)	
C16: It's very strange. Feeling this feeling underneath it, and trying to talk, right now.	C16 She describes having a felt sense. She finds it odd. There is the presence of "this feeling" which is "underneath," so that she cannot very well talk without losing hold of it. She makes it clear also that there are no words to talk from it. It is an unclear, single "this."
T16: Sensing the feeling directly and trying to say what it is. And it's scared.	
C17: It's very interesting, the fear is right underneath it. Now I'm content to just sit there with the withdrawn, and feel apathy until I … end up with the feeling, then I withdraw into the nice apathy again. (*Laughs*)	C17 She describes going back and forth, sensing the "scared," pulling back into apathy. She calls it "the nice apathy" and laughs because, of course, she doesn't really want the apathy, but she can sense directly how it is more comfortable.
T17: Mhm, the apathy is more comfortable and the fear is right under it, so you just push down and ah … there it is. (*Silence*)	

Transcript	Gendlin's first commentary
T18: Well, let's be friendly with the fear, and sort of say, that's all right, right now we're not doing anything. We'd just like to hear from it. What it's so scared about. (*Silence -- 3 minutes*)	
C19: This is an all-good part of me but it would rather be dead, than come out to … um … being tromped on.	C19 Now the shift has taken place. Something new has opened, and it turns out that this is an "all-good" part of her, that pulls back. She senses the reason for the pulling back from *the inside of* the pulling back, or more exactly, from inside this newly sensed "all-good part of her. It would rather be dead than be received like that, but she senses this part of her (that seems much more significant than simply being the reason for the pulling back).
T19: It's all good, but if it's going to get tromped on, it would rather be dead, or stay pulled back. (*Silence*)	
T20: Now, can you really be glad that part came out and that it's speaking to us? Can you welcome it?	
C21: It's like … when you're just being nice to a person, and someone watching later tells you that you were just trying to buy that person.	C21 She describes why it pulls back. The good of this part is received by others as bad. She gives the flavor of that in an example. She was nice to someone. Another person watching accused her later of "trying to buy that person."
T21: Inside you it's good and then they make something bad out of it?	
C22: Yes. (*Silence*)	
C23: Well, that sure is different.	C23 She means that her pulling back has turned out to be something very different than she had expected, and said earlier. Rather than being negative "what pulls back" has turned out to be a good and loving part of her.

Gendlin spends two entire chapters discussing the interactive "focusing" process taking place in this excerpt. He lays out a theoretical construct for noticing eight characteristics of a “process step,” and then details six kinds of therapist responses that can help engender process steps. His technical terms are intensely colloquial; it is very difficult to grasp them outside a lived human-human dialogic situation because they are so pragmatically structured. Even in accepting the full context of Gendlin's written argument, one has to put Gendlin’s terms into practice in one's own experience to recover their nuances. I want to assert, furthermore, that all professional psychotherapeutic discourse, regardless of its personality theory, requires some direct association with practicing therapy in order to attain its full significance.[36] I mention all this to make the framework of practices that accompany actual psychotherapy explicit, as it is often unconsidered or unknown in popular or even scientific reference to the discipline.

We will restrict ourselves to using a few of Gendlin’s terms within arguments so that the terms might at least carry a minimal semantics, full well recognizing the entire loss of pragmatics. To attempt to blankly define these terms in a "context-free"[37] manner would be the worst sort of reductiveness: much better to go to the original source. In comparing the ELIZA productions and the Gendlin transcript, we will be able to touch on only a small subset of the richly hermeneutic framework that Gendlin.

Demeaning Features of ELIZA-Person Dialog

1) Semantic Abuse

ELIZA was designed with mechanisms to make sure (she) always has something to say. Thus no matter what is input, or even with no input at all beyond a carriage-

---

[36] I was fortunate enough to learn Gendlin's framework in a clinical psychology practicum, and applied it for a period as an in-patient child care worker for seriously disturbed children at the Hines Veterans Administration Hospital in Palatine, Illinois.

[37] See footnote 47, p. 55 below for a further argument about the significance of this term.

return, E always generates a response. These mechanisms generate a powerful illusion of semantically meaningful interaction. E recurrently throws up some symbols calculated to have "semantic pull" regardless of whether they add any relevant meaning to the interaction. The program does not just happen to trip us up with our natural language propensities for projecting anthropomorphic meaning; rather it **depends** upon our inevitable anthropomorphism for its continued appeal. It is not generating autonomous meaning with human-human parity. That there might be highly relevant, important symbolizations achieved through calculation -- as is the case with computer modeling, data reduction, and other use of computer power as a tool for symbolic manipulation – I neither deny nor want to deal with here. My point is that even in such profitable cases, the success of the human-computer interaction is fully dependent upon the person's projection of meaningfulness upon the symbolic artifacts. That the machine does its calculations on "its own" only increases the effect of its imitation of symbolic autonomy.

When the person's own semantic dynamic becomes tied to E's responses, the dynamic quickly decays into a semantic battle. E's constant return with "yet another tantalizing or frustrating response" is an instance of what I term *semantic abuse* or *dictive intimidation*. Such logomachy is the first of four modes of "meaning loss" I will identify.

2) *Goal Thinning*

One of the deep features of a therapeutic interaction, according to Gendlin, is what he calls a "felt shift" in the experienced meanings from which a client is responding to the therapist. Fostering such changes in the client's experiencing is a definite goal of the therapeutic discipline.

The example of a felt shift in our transcript emerges in the sequence starting at C13 and going through C23. Let’s refer to the transcript again.

> • at C13 the client acknowledges that the therapist's reflection at T12 closely matches "what the client means at that moment."
>
> • from C14 to C17 she stays with and explores the nuances of her "felt sense," while the therapist's responses at T16, T17, and T18 help her to keep in touch with these normally overlooked aspects.
>
> • then from C19 to C21 new meanings are explicated, that is, given unprecedented symbolization that comes out of the felt sense.
>
> • at C22 and C23 the client gives her own autonomous confirmation that her inner world of meanings is genuinely different.

This shift in the client's experienced meanings is what Gendlin calls "content mutation" and is precisely the sort of meaning creation that cannot be reduced to mere symbolic manipulation, much as logical formalisms do not properly express metaphors.

In a world confined to the client-therapist interaction, these felt shifts amount to a deeply significant human communication that parallels learning, acting toward a end, and creative thinking in the wider, less sympathetically structured world of everyday experience. In those wider contexts, we experience "felt shifts" as insights, or accomplishments, or interactive satisfactions such as the famous "Aha!" experience.

Conversations with ELIZA lack altogether or realize only incompletely these kinds of satisfactions, in systematic ways. The one mechanism that E possesses which might even approximate such a felt shift is that of encoding some small information about the "real world":

> It is very often true that when a person speaks in terms of universals such as "everybody", "always" and "nobody" he is really referring to some quite specific event or person.
>
> …
>
> For example, the transformation rules which cause the input

> Everybody hates me
>
> to be transformed to
>
> Can you think of anyone in particular
>
> and other such are based on quite specific hypotheses about the world. The whole script [for ELIZA's transformations] constitutes, in a loose way, a model of certain aspects of the world. The act of writing a script is a kind of programming act and has all the advantages of programming, most particularly that it clearly shows where the programmer's understanding and command of his subject leaves off.[38]

Thus, along strong AI lines, the psychologically astute programmer might attempt to build in all sorts of beneficial "psych-outs" to make E into a real therapist. Inevitably, this kind of programming is faced with the attempt to encode a vast array of extremely context-sensitive knowledge and skillful interactive behaviors as exhibited by a human therapist in terms of AI syntactic and vocabulary lookup mechanisms.[39] We can readily imagine additional instances of semantic abuse along the lines of ELIZA running through "transformation" after "transformation" hunting for a catchphrase that will strike some emotional chord with the client. But hidden within that possibility is another type of loss, namely that of diminution of beneficial "felt shifts." Even should E hit upon an appropriate conversational twist, it would be at best only marginally able to follow up on it along with the client so that the client might autonomously "take it in for

---

[38] Weizenbaum, "*ELIZA*," pp. 39 and 43.

[39] This situation is a classic example of an error that Dreyfus attributes to the "metaphysical assumption": "More recent work has thus been forced to deal directly with the background of commonsense know-how which guides our changing sense of what counts as the relevant facts. Faced with this necessity researchers have implicitly tried to treat the broadest context or background as an object with its own set of preselected descriptive features. This assumption, that the background can be treated as just another object to be represented in the same sort of structured description in which everyday objects are represented, is essential to our whole philosophical tradition. … My thesis, which owes a lot to Wittgenstein, is that whenever human behavior is analyzed in terms of rules, these rules must always contain a *ceteris paribus* condition, i.e., they apply 'everything else being equal,' and what 'everything else' and 'equal' means in any specific situation can never be fully spelled out without a regress. Moreover, this *ceteris paribus* condition is not merely an annoyance which shows that the analysis is not yet complete and might be what Husserl called an 'infinite

herself." Such a possibility assumes some master script that is much, much more artful at very crude input analysis, an assumption which is itself extremely dubious. In this fashion the therapeutic goals of engendering beneficial felt shifts are reduced or significantly <u>thinned out</u> in client-ELIZA interactions. More generally, human interactive goals are very likely to be poorly realized in human-computer communications, except to the extent that the user views and understands the program as merely a tool, an artifact useful for some human purpose. People working with computers without such attitudes will implicitly be subject to *goal thinning* or *purposive truncation*. This is a second mode of meaning loss.

3) *Surplus Meaning Deprivation*

One of the most difficult dimensions to capture of a genuine client-centered interaction is precisely the aspect of a "reflective technique" that makes this interaction so readily open to caricature. Outside of the real therapeutic setting, that is, one in which a client is genuinely disturbed about some inner meanings that are somehow inaccessible to change, the reflective responses arising out of Rogers's principle of "unconditional positive regard" seem all too platitudinous in "normal" life. This facet is what Weizenbaum mapped to "content-free remarks" that would capitalize on the user's willingness to assume that the speaker "had some purpose in so directing the subsequent conversation."[40] It was to be the rhetorical device that would conceal just how little real-world knowledge ELIZA had. The mistake here is terming such remarks as "content-free" ignores the disciplinary framework of psychotherapeutic practice with its extremely context-sensitive pointers or deictic markers for the dialogic process. Exactly at the point where AI is looking for abstract, brutally syntactic features to

---

task'" (Hubert L. Dreyfus, *What Computers **Still** Can't Do: A Critique of Artificial Reason* [Cambridge, Mass: MIT Press, 1992], pp. 56-57.

manipulate, humans are at their most pragmatically nuanced best. We can see something of what an actual therapist is doing to adapt to the client's meaning process in the following recapitulation of a part of the session excerpted above with an added second layer of Gendlin's commentary from the therapist's standpoint. The left-hand column is the same as before, while the right-hand column now gives a second commentary that reviews the therapist's actions:

---

[40] Weizenbaum, "*ELIZA*," pp. 37, 43.

	Transcript	Gendlin's second commentary
T5:	And "jittery" is the best word for it.	He thinks "jittery" is the quality-word that would let her sense the whole of this.
C6:	Yeah, yeah. Uh … I pull back.	
T6:	Pull back is it.	Although she says "yeah," the "Uh … I pull back" lets him know that "pull back" gets a hold of the whole of this feeling more deeply.
C7:	The jittery is more a surface than the pull back. The jittery comes when part of me says, "well, you know, you really have to do it now."	
T7:	So we don't know what pulls back, it's not the jittery that pulls.	He would like her to sense that, right there, which wants to pull back.
C8:	No, the jittery is a result.	She doesn't do what he hoped she would do.
T8:	So we don't really know what the pull back feels like, what it is that wants to pull back.	He tries again to help her let the "what wants to pull back" form right here, so she can sense it. If she sensed it, that would provide a fresh opening.
C9:	Well I think it's … ah … that I don't want to test myself. And I'm afraid, ah, the bad things will be confirmed.	Instead, she tells how she thinks about the problem.
T9:	Can you feel the pull back, if you imagine yourself going ahead?	He ignores what she says. He invites her to imagine going ahead, so that the pulling back will come to her right now, concretely.
C10:	Yeah, I can feel the pull back now. … The pull back is into weed, that's what it does.	
T10:	Into weed.	He does not understand what she means.
C11:	Marijuana, that's the perfect place to pull back.	She explains that she can now feel how smoking marijuana is, for her, a perfect way of pulling back from living and from situations. The role of marijuana in her withdrawing becomes clear to her, if she had not known it before. She finds it here: the pulling back is into a drug.
T11:	That's a perfect place to pull back to.	

Transcript

C12: Yeah. But if I don't go to the line then I don't have to pull back.

T12: As long as you don't really go across the line, there is no testing of it, there is no proof, good, bad, and you're suspecting that you're afraid of actually finding out.

C13: Right!

Gendlin's second commentary

C12: She reiterates what she has been trying to say.

T12: At last he responds exactly to her own analysis of the problem.

C13: She gives him a strong "right!" She feels this time (at last) he took in what she thinks about it.[41]

---

[41] Gendlin, *Focusing-Oriented Psychotherapy,* pp. 30-32.

What we see here is a later reflection on the therapist's exercise of his skills in the actual interaction as it was transcribed. That second-order systematic reflection and training go to make up and revise the interpretive discipline at the opposite end of the spectrum from the programmer's efforts to redesign ELIZA's word transformations. Gendlin identifies this kind of close tracking as a special art:

> Reflecting, or "listening" (as we call it) includes saying back exactly what the person is trying to convey. The therapist attempts to grasp *exactly* how each bit of her experience, moment by moment, feels to her. He wants to be in contact with every turn she takes, with every one of the meanings she finds, as they are to her. This is the therapist's intention. Of course every therapist will misunderstand at times, and will inadvertently ignore some messages for a time. As soon as the therapist realizes this has happened, a fresh effort to understand has first priority. Whatever else the therapist may do is never done without recognition of exactly what the individual is intending to convey.[42]

The real-time therapeutic interaction with its uncertainties, surprising twists and turns, and requirements for metaphoric creation, along with the reflective personality discipline, are to be contrasted with what Weizenbaum refers to as "the act of writing a script" which he describes as a "kind of programming and has all the advantages of programming, most particularly that it clearly shows where the programmer's understanding and command of his subject leaves off." There is a palpable gap between the two disciplines in the many sorts difficulties entailed. A programmer who could successfully pull off getting the correct restatement of each of the client's meanings as they occur in "real-time" would be akin to a Sophocles having writing – or "programming" – a separate play for each Greek citizen to ensure that they would personally achieve a catharsis of their pent-up emotions. But even that would be orders of magnitude of complexity away from the nuance and subtlety of the client-therapist

---

[42] Gendlin, *Focusing-Oriented Psychotherapy,* p. 45.

interaction, because the typicality of a culture's unique sense of tragedy would not come close to the diverse range of problems faced by individuals. Even the live therapist has no prior knowledge in detail of just how the client will come to experience personal difficulties, despite the possibility for diagnostic categorization with a high degree of accuracy.

What this unpacking of orders of difficulty brings out is the immense possibility for surplus meaning and overdetermination in human-human interactions – and a corresponding third mode of meaning loss. It is in this richness of meaningfully nuanced subjectivity that people experience joy and triumph, failure and hurt. To convert the possibilities for symbolic and gestural nuance into "content-free remarks" is a kind of *surplus meaning deprivation* or *insignificant precision* of programming constructs without a possibility for accurate application.

Although it highly unlikely that any programming achievement will realize such capabilities with direct relevance to individual needs, the problem does set something of a "requirements definition" for the new dramatic arts of a genuinely moving "interactive fiction." Presuming the participant's willingness to enter the world of the author's digital art work, and an ability to suspend one's own sense of reality within the virtual world, the culturally enriching rewards of new genres may be singularly compelling as virtual realities. What is clear, though, is that the author's of such art works will be human, not computers.

4) ***Communicative Channeling***

Another dimension of a genuinely therapeutic encounter is its orientation toward opening the client's self to new behaviors in the social world outside the therapy hour. Our very descent into psychological complexity marks the need for a persistence of change outside the client-therapist interaction. A successful outcome has to see the client move from testimonials of "only with you (the therapist) can I feel really myself" to a more independent and growth-oriented stance of being "up to the challenges of life." Indeed, much of the controversy about psychotherapy stems from the great difficulty of securing exactly that transition with any great predictability. The exemplary success of our relatively in-touch client lies in her declaration, "Well, that sure is different." Therapists do not always engender such change steps, nor do the steps always carry over to function in the client's other relationships.

However overtly sympathetic and supportive a therapist might be, that feeling quality in itself is insufficient to foster change and a fresh opening to life's possibilities. Just as crucial is further inward work on the client's part that the therapist intends to collaborate in, provided he or she can move beyond sympathy into an empathic response that can engender healthy restructuring by the client. In a fashion the therapist implicitly elicits healthy plot lines or schemas with which the client can go about working with their own felt sense of a problem. Such implicit schemes allow the client to create highly specific metaphors and other creative tropes specifically adapted to the particular felt referents of the client, which are sometimes powerful enough to engender significantly different interactions both in and outside of the therapeutic relationship.

For example, in our transcript we see the client doing exactly that in C14 through C23 as she follows the structuring moves, first, to let a felt sense of "pull back" settle out, next to let a new quality of "scared" emerge "underneath" the "pull back," find a comfortable "apathy" still on top of the fear and, then, with the therapist's guidance to "just push down" and "be friendly with the fear" so as to engender a response directly from a narratively displaced "fear" that can now speak for itself. The client then moves directly into reporting an "all-good part" of herself that is genuinely different from her initial "I don't seem to think a hell of a lot of myself."

In this sequence, the therapist is not <u>telling</u> the client anything about her own felt meanings. This is not an occasion for pronouncing a psychiatric interpretation that is "right" about the client independent of her acceptance or resistance. Rather, quite differently the therapist is making narrative space for the client in which she is able to find her own metaphors and create new relationships to her felt referent.

Let's compare that process of making meaning to an ELIZA script. All things are not equal here,[43] the intrusion of differences is intrinsic to anything approaching full human complexity. So, with that in mind, let's begin by looking at the tough case (for anthropomorphic humans), where the real client's responses were fed into ELIZA regardless of ELIZA's responses:

C: Yeah, I could feel it.
1E: You say you could feel it?
C: Scared … it's like the world is going to bite me or something.
2E: I am not sure I understand you fully.
C: It's very strange.
3E: Please go on.
C: Feeling this feeling underneath it, and trying to talk right now.
4E: What does that suggest to you?
C: It's very interesting, the fear is right underneath it.
5E: Do you feel strongly about discussing such things?
C: Now I'm content to just sit there with the withdrawn, and feel apathy.
6E: Is it because you are content to just sit there with the withdrawn that you came to me?
C: I end up with the feeling, then I withdraw into the nice apathy again.
7E: Can you elaborate on that?
C: This is an all-good part of me but it would rather be dead, than come out to being tromped on.
8E: That is interesting. Please continue.
C: It's like when you're just being nice to a person, and someone watching later tells you that you are just trying to buy that person.
9E: What makes you think I am just being nice to a person?

---

[43] See footnote 39 above for Dreyfus on *ceteris paribus*.

C: Yes.
10E: You seem to be quite positive.
C: Well, that sure is different.
11E: How is it different?

First off, it is a tribute to Weizenbaum's scripting art that ELIZA comes back with six different "content-free" responses (E2, 3, 4, 5, 7, & 8) that at least don't get in the client's way. Of course, this is because the program's limited vocabulary of fifty key words simply fails to recognize most of the metaphoric content and therefore defaults to empty platitudes. Nevertheless it makes for routine reading. 1E and 6E actually work and could be considered decent stepwise reflections – but, again, these are just the simple swapping of keywords by template. On the other hand, 9E, 10E, and 11E fail outright by revealing definite semantic gaps. In sum, ELIZA does a passable job of keeping up with the client's steps as engendered by a human therapist. (She) just might be able to hide out in a group of therapists without immediate detection – a kind of professional conference-style Turing test, I guess.

Not surprisingly, this is a tribute to exactly what can be useful about computerized symbolic manipulations. Given a real process or real data, they can often perform calculations that achieve symbolic results of legitimate interest to people (or, in this case, at least not immediately alienating). In a strictly didactic context, a human-computer dialog can definitely be constructed around some subject matter content which a user is trying to learn or an transaction he is attempting conduct. While linguistically wooden, such interactions can indeed be educative or instrumental.

Now let’s go back to a previous transcript to look at the tough case for computers trying to lead people:

1b) ELIZA: How do you do. Please tell my your problem.
Client: I was thinking about - on my way over - I don't seem to think a hell of a lot of myself.

> E: Your way over you don't seem to think a hell of a lot of yourself?
> C: I'm not way over myself, I was just not thinking a hell of a lot of myself.
> E: Is it because you are not way over yourself that you came to me?
> S: It really isn't that bad - sure I have low self esteem, but I'm not way over myself.
> E: You say it really isn't that bad sure you have low self esteem?
> S: Yeah, I don't seem to think a hell of a lot of myself.
> E: Don't you really seem to think a hell of a lot of yourself?
> S: No I don't. You are making that up.
> E: Are you saying no just to be negative?
> S: No, it's because you are insulting me.
> E: You are being a bit negative.
> S: No I'm not.
> E: Why not?

Here we see again the degenerative pattern of miscommunication noted above. Rather than providing healthy plot lines or schemas for the client, ELIZA only proffers quite distracting instances of a failing interaction – not particularly therapeutic.

Hence, we could conclude that ELIZA is routinely passable, sort of a beginner in a class of professionals, a novice that with time and artful scripting might just make the grade. But this conclusion would be valid only under the metaphysical assumption of "all things being equal." We are most definitely not in an equal situation. What we have been following here, as readers, is a humanly written discourse about a symbolic artifact, a computer program. We are already acculturated to accept such symbolizing as if it were a direct human-human interaction. But that too requires an immense suspension of disbelief. It takes years of collegiate education to learn to get the substance of the most routinely presented fact-filled textbook, much less plumb the depths of a great work. What is at stake here is the "felt sense" of the lived experiential qualities that each of us has and is throughout our lives. Computer symbolic manipulations do not arise from felt meanings. There is a vast chasm between Panbanisha's "Why was she being mean to him" and the most sophisticated mathematical products of a program. This capacity is what Dennett would have us discard as some mystical folk theory of consciousness, while he substitutes for it with an abstract symbolic account of the brain or whatever.[44]

ELIZA's conversational gambits appear as marginally reasonable drivel only because of their encapsulation in a wider discourse. ELIZA is in no manner a great communicator and, thus, is profoundly inadequate to foster pragmatic restructuring – much more so than any written account can communicate. In fact, both Weizenbaum and Dreyfus remark on the strange lack of persuasiveness their analyses carry in the face of boostering on behalf of AI.[45]

The range of schematic or topical restructuring between marginally educative exchanges and semantic failure that we have thus far explored, points to a general property of human-computer interactions that I call *communicative channeling* or *topical foreclosure.* Even when the human-computer interaction works, it does so more in terms of the preprogrammed content or activities that the program presents to the user than in the user's natural language terms and everyday behavioral processes. This, then, constitutes a fourth mode of meaning loss.

If we grant that the practice of psychotherapy is more an art than a science, and dependent on many imponderables about both the client and the therapist, should we

---

[44] The difference between Dennett's view and the one presented here is that "qualia" are not taken as some dead-end perceptual experiences (i.e., the raw feel of "red" or "pain" or "snake-yuckiness" (p. 385) or whatever) that clearly cannot carry much further significance beyond their experience as such. The view here is that a "felt sense" is the experienced awareness or vehicle for all of our experienced meanings, most especially the ones that are pointed to by Frege's use of "sense" as distinct from "reference" for propositions. Only here, the meaningful sense of an expression is not taken as fixed in the service of logical argument. Once that constraint is relaxed, it is possible to inwardly discern qualitatively different nuances of meaning without reducing them to raw percepts. Those different nuances of meaning are given together in our awareness and, are capable of further differentiation. Not only is formal logic not particularly helpful in the psychotherapeutic context, it is of limited usefulness in the rough and tumble context of state-of-the-art science. Only after desired, but incompletely realized, meanings have been disclosed do formalizing techniques play a powerful role.

[45] Dreyfus, *What Computers **Still** Can't Do*, p. xi ff. Weizenbaum, *Computer Power and Human Reason*, ch. 10.

abandon it as only a pseudo-science, capable of interpreting any hypothesis without constraint to consequences, as Popper implies? And instead, should we gladly acknowledge the more structured framework of artificial intelligence as superior? The problem here is not whether or not the AI artifacts will ever be up to the work of intending genuine restructuring: it is extremely unlikely they ever will. Moreover, Weizenbaum argues overtly against the morality of this possibility – he labels it an "imperialism of instrumental reason."[46] The problem is how to reconceptualize the neo-Cartesian metaphor of "mind as automaton" into a context in which computerized symbolic manipulations are free from unwarranted or misdirected anthropomorphisms. Viewing them fundamentally as artifacts is a step in that direction.

Anthropomorphism and Anthropocentrism /Technical Deepening and Symbolic Demeaning

Not only do terms carry a history of meanings and uses; they also can undergo an extensive development of conceptual or disciplinary significance by when they are utilized as new technical terms for a science or art. Thus far my inquiry has been into contemporary applications of the "mind as computer" metaphor through a detailed analyses of anthropomorphism in human-computer interactions. I do believe that large numbers of people using well-designed programs can bring us all to a cultural watershed. So, let us put these studies into the background and begin anew, as Aristotle might say, so as to enable ourselves to gain further insight into our possibilities for being human in an age of computerized artifacts.

---

[46] Weizenbaum, *Computer Power and Human Reason*, p. 269.

Starting afresh with a historically empty, or relatively "context-open"[47] semantics for our two key terms, we can take anthropomorphism to mean, in ordinary language, "the projection of human traits onto the nonhuman," and anthropocentrism to mean, also in ordinary language, that "all nature is constructed for human ends."

To date, in our contemporary dogmatic assertion of anti-anthropomorphic methodologies, we have turned out to be either explicitly or unwittingly anthropocentric. Goodall, Jolly, and others have recognized this tendency with regard to animal rights, but the problem has an even deeper aspect. We can readily formulate Goodall's and Jolly's line of argument as an anti-anthropocentric moral stance. Focusing on the anti-anthropocentric intent of their stance is also all too convenient for letting ourselves off the hook. We tend to be falsely or unreflectively anthropocentric

---

[47] It may seem inconsistent to use "context-open" in a positive fashion after denying the appropriateness of the Strong AI reduction of natural language phenomena to "context-free" terms, as I do at page 65, (AI's terms as inherently context-free because they lack a felt sense) and refusing to provide "context-free" definitions of Gendlin's terms at p. 81, (since Gendlin's terms are strongly context-dependent). In fact, the situational difference between a typical AI reduction of all the rich aspects of natural language use that we saw exemplified in ELIZA, versus the philosophical trope of suspending the historical semantics of base terms in service of creating new meanings is quite striking, if often unnoticed. The former denies the richness and variety of meanings because they are inconvenient and messy for logical formalisms or, in our case, artificial language constructs, whereas the latter begins by recognizing the history of terms and their "use families" as changing and evolving "forms of life" and then creates a narrative separation so as to allow the emergence of new meanings and applications for the root semantics. This latter "context-open" approach to the diversities of natural language is strictly nonreductive. Rather than believing that words can have their meanings abstracted and fixed for all time in a realm of formal reduction, it asserts that words and language are our best and most flexible means of discovering the features of the world around us, and furthermore, that our linguistic interactions are constant sources of meaningful change and stability. It eschews the naveté of generation after generation of AI projects, each of which will find "the ultimate" reduction of meanings to a fixed set of primitives in favor of the longer term realization that specialized language use requires socially and technically stable usage so long as there is active research and fresh discovery taking place in some program of inquiry. After that they become historically situated semantics. Finally, in the intensely situational communications of psychotherapeutic interactions we have an especially strong pragmatic context in which the very meanings of words and the very forms of expression are often in fundamental flux: a veritable microcosm of meanings in transformation. This is something to be experienced as almost "all-context" rather than defined as "context-free". Of course, from the Strong AI standpoint such highly pragmatic utterances appear not as "context-dependent" but as "content-free" because they appear to lack propositional content.

in our use of methodologically impersonal technological metaphors. Explaining human cognition as identical with computational mechanisms is a primary example of this practice. The objectivity of description obtained by such neobehaviorist methodologies comes at the cost of obscuring our instrumental relationships to the life forms studied, which in this case are people. In our anti-anthropomorphic and anti-anthropocentric stances toward other creatures we have let ourselves off the hook, so to speak, by excluding the natural place of humanity's own productive sciences from reflective analysis. We have lost Aristotle's balance:

> But as in human operations, so in natural processes; and as in processes, so in human operations (unless something interferes). … In general, moreover, art completes what nature is unable to carry to a finish; or art imitates nature. If, then, processes by art are to some end, it is clear that natural processes are too.[48]

We now need to reconstruct not the rest of nature to human ends (a project all too well in progress) but, rather, our own artifacts as belonging to our natural productivity. To this end, we must become rigorously anthropocentric toward our artifacts as well as morally anti-anthropocentric toward other creatures. Similarly, we need to understand and appreciate the intrinsically anthropomorphic character of our primary artifact, natural language, and how that projective characteristic is severely absent from our more recent invention, artificial language. To this end, we must seek designs for artificial symbol systems that are more natural and supportive of more ideal communications.

Placing the above relatively empty semantic structures for our two key terms into the context of our contemporary problematic for human-computer interactions, we can discern new directions opening for their meanings. Our central problem is that no one

---

[48] Aristotle, *Physics*, ii.8 - 199a10 & 17.

can long resist the impulse to attribute human meanings to the productions of artificial symbol systems. At the beginning of this paper, our emblem of disclosing unconscious anthropomorphism was Gary Kasparov's encounter with Deep Blue. Kasparov’s shift in attitude from his first attributing human strategies to Deep Blue to his later discovering of the artifactual character of his simulated opponent as a manipulator of a brute force algorithm provides us with an exemplary case for a desired "user" transition from the unwittingly anthropomorphic to the artifactually sophisticated:

> I got my first glimpse of artificial intelligence on Feb. 10, 1996, at 4:45 p. m. EST, when in the first game of my match with Deep Blue, the computer nudged a pawn forward to a square where it could easily be captured. It was a wonderfully and extremely human move. … Later I discovered the truth. Deep Blue's computational powers were so great that it did in fact calculate every possible move all the way to the actual recovery of the pawn six moves later. The computer didn't view the pawn sacrifice as a sacrifice at all. So the question is, if the computer makes the same move that I would make for completely different reasons, has it made an "intelligent" move?

We can see now that the answer to Kasparov's question has to be, "That is the wrong way to frame the question." If we look closely at Kasparov's sequence of actions "to see what moves" (to use Dennett's phrase), we can notice an iterative cycle of anthropomorphic interpretation and continued play of the game, followed by an "offline" learning and reflective questioning preparatory to another game. In this fashion we can cultivate a progressive understanding of the status of computerized artifacts as "objects of symbolic design" (that is, how they encode features of reality in data structures), learn to search for their relatively context-free mechanisms of operation (that is, how they execute algorithmic and heuristic procedures), and seek to determine the constraints on their modes of interaction and communication (that is, how they structure input and output and how they organize their symbolic representations of internal states for user interface). Wise in the ways of computers, the user could then refer directly to his or her felt sense of "how humans would do it" as a proper means of shaping the use of (or winning against) artificially intelligent symbol systems.

That series of steps is how Kasparov as chess player – or, more generally, as computer user – can properly cycle through his human-computer interactions. From the other side, the IBM scientists are faced with a different cycle, the cycle of design. They are focused on reprogramming Deep Blue, seeking ever more impressive symbolic manipulations with the appearance of intelligence. They have a fascinating medium in which to work. Artifactual symbol systems generate algorithmically and heuristically structured meanings for us with unprecedented precision and stability. Never before have our symbolic artifacts consisted of mechanisms capable of automatically manipulating symbols as if they were physical objects. Never before have we had occasion to observe and interact with symbolizations presented by artifacts external to ourselves and lacking the intrinsic anthropomorphism of natural language. All of these aspects of artifactual symbol systems provide the opportunity to designers for a "technical deepening" of our arts and sciences. The introduction of the printing press, the construction of machine tools in the Industrial Revolution, and the introduction of the transistor into computers are examples of technical deepenings. Each can have dramatic and widespread results in the capabilities, uses, ease of use, problems solved, and other properties of a technological capability. Such differences are not predictable from the basis of the prior generation, and in that sense a genuine technological revolution takes place. All of these factors and characteristics make design cycles dynamic and replete with opportunities.

If we thematize the technology as encompassed within the context of humanly productive science, we are in consequence afforded two primary viewpoints. Kasparov's viewpoint is that of the user; users may benefit from use and play, subject all the while to possible interactional pathologies, or "symbolic demeaning." The IBM scientists who created Deep Blue worked from the viewpoint of the designer: artifactual symbol-system designers may achieve technical deepenings with multiple productivities, responsible all the while for the interactional properties of their constructions. The merger of these two viewpoints, user and designer, combines humanistic arts with their reflexive access to experiential nuance and computational arts with their formal precisions in a larger, nascent whole. Let's see if we can reconstruct the semantics of anthropomorphism and anthropocentrism through the use of these two viewpoints.

Anthropomorphism in human-human, natural language interactions is unavoidable and desirable. We inevitably project our meanings and traits onto each other as well as onto our natural language artifacts. The latter is especially the case for those linguistic artifacts with great human centrality and semantic intricacy – that is, the great works of our cultures. This projection serves us well as a discipline for detecting and understanding symbolic nuance as we attempt to recover the rich worlds of human experience codified in our texts and conversations. Yet, such a hermeneutic anthropomorphism stands in need of correctives supplied by methodologies of mock anthropomorphism applied to the study of all natural objects and may benefit from technical deepenings to the hermeneutic arts as they are being significantly extended by access to artifactual symbol systems.

A primary claim I am making in this paper is that anthropomorphism in human-computer, artificial language interactions is unavoidable and undesirable in its naive form. Projected with all the inevitability it acquires in natural language interactions onto artificial symbol systems lacking the same communicative richness, anthropomorphism encounters dangers of systematic loss of meaning. As we have seen, two attendant dangers arising from naive anthropomorphism are, as I term them, communicative channeling and goal thinning.

Natural language is metaphorically projective. But because artificial symbol systems cannot evolve through metaphor and other creative tropes, they present fundamentally closed or rigidly channeled communications. Everyone has had the experience of trying to get a computer to do something slightly different than what it is designed for and being utterly frustrated with the "closed-mindedness" of the interaction. This loss of meaning may be addressed through an explicit design goal of fostering *enriched experiences of symbolic agency* wherein the user's activities are better represented within the program and likely patterns of interaction are anticipated. Brenda Laurel's concept of "computers as theatre" is discussed below as an example of a designer working toward such technical deepenings.

Natural language is primarily expressive of human interests and purposes. Because artificial symbol systems can define achieving satisfaction only in truth-functional or concrete-operational terms, they can develop and respond only to digital implication networks which are obviously less robust and much more brittle than human goal-oriented behaviors. Humans, on the other hand, can adapt to complex circumstances and shift objectives in "satisficing" ways that are much, much richer and more likely to achieve at least partial satisfaction in a more continuous fashion. While computers are immensely useful as tools, people nevertheless are forced to truncate their purposes in human-computer interaction. This loss of meaning may be addressed through an explicit design goal of searching for greater *theoretical completeness* and robustness of interactions. Unfortunately, this goal presents a difficult task that is only now receiving direct attention. We will have to be satisfied with a report of "routine practices" in the next section.

A second primary claim is that anthropocentrism in human-computer, artificial language interactions is highly desirable and has been systematically underemphasized or forgotten by many design communities and advocates of the Strong Program. This omission has occurred because of their failures to recognize the possibilities for pathological interactions and their tendencies to categorize computational models of consciousness and intelligence as equivalents to human functioning rather than as artifacts. Much of my previous discussion in Part 2 attempted to support this assertion. Yet the technology is in tremendous innovative flux. Technical deepenings of all sorts are open to selection and development. With a different conceptual orientation toward productive science and with a better understanding of how our "human as computer" metaphors mislead us, it is quite feasible to technically deepen aspects of symbol systems in support of our human natures. Thus, two design goals that arise from a reconstructed anthropocentrism are, as I term them, *greater conversational expressiveness* and *promoting cross-coherent symbolic worlds*.

Artificial language relies on strict syntactical and structural relationships for logically constructing semantic connections. In human-computer interactions, artificial symbols can be extremely powerful tools for manipulating concepts as "intuition pumps." They can also discredit, ignore, or warp legitimate meanings present at the semantic level in the person interacting with the computer. Moreover, this general characteristic of artificial symbol systems is often exacerbated by implementations with strictly limited vocabularies that, in effect punish, people for simply using different words than the ones chosen for the symbol system. Such semantic abuse can be addressed by adopting a proactive design goal of *greater conversational expressiveness*. Guidance for designers can be found in Grice's cooperative maxims[49] but his maxims are themselves more oriented towards constraining human discourse to forms more amenable for programming. Additional arts are needed.

Artificial symbol systems are radically dependent on univocal significance from their human interlocutors. Programs must at least initially remove or exclude all surplus of meanings and even redundancy of expression and action from their input. This allows for very precise and unambiguous communication. Human-human interactions, however, are very much based on the presence of multiple levels of redundancy. People interacting with artificial symbol systems are not given the subtle cues for shaping communications into meanings that are normal in human-human interactions and therefore are forced to make up for this loss from their own experiencing. Consequently, the absence of such multiplicity subjects humans in human-computer interactions to conditions that are inherently stressful. Over extended periods of time, such demands for denotative precision without nuanced expression lead to surplus meaning deprivation. This problem can be addressed in part by promoting links to additional meanings and a multiplicity of interactive contexts. In general this can be formulated as a design goal of promoting *cross-coherent symbolic worlds* within human-computer interactions. I will explore this goal in terms of the recurrent design problems of hypertext in another paper.

Anthropocentrism in human-human, natural language interactions again seems unavoidable for better or worse since the full range of human behavior takes place in this medium. It has been subject to a variety of religious, legal, moral, economic, political, and other human scientific analyses which at times seek to improve the

---

[49] Paul Grice, *Studies in the Way of Words* (Cambridge, Mass. : Harvard University Press, 1989.)

character of human agents and at times seek to ameliorate damages done. Yet natural language is under confusing or paradoxical stress through the introduction of artifactual symbol systems in ways that put much of our prior understandings at risk or, at least, subject them to great change. It is this pressure that motivates me to focus on articulating a moral economy of symbol system design.

Steps toward a Moral Economy of Design

The third primary claim I make here is that artificial language constructs in all the multimedia modes of digital symbolization cannot long be kept separate from traditional natural language discourse and our embodied meanings in all their kinesthetic, visual, and auditory modes of interaction. Demographically, millions of people are already having "conversations" with computers on a regular basis, even while they continue to converse with each other. Our natural communicative skills are far too adaptable to reject or miss all the symbolic innovation and structural richness of artifactual symbol systems. Not only are we inventing all sorts of technical terms for computers, we are also absorbing computerized symbolic structures into our range of person-to-person discourse. An impressive, but simple, example of this was the emergent practice starting in the late 1980s of publishing or distributing "live" documents in Mathematica™ Notebook form. Malcolm Slaney was one of the first to publish a Mathematica™ article. His work on Lyon's Cochlear Model contained both the text of his article and the models including the computational structures for exploring them. This combination allowed readers to alter the parameters and provided them with the capability for checking as well as extending Slaney's thought in the same computational environment used by the author.[50] Further discourse could then be conducted fully in an operative computational framework. Such capabilities are fast becoming routine on the World Wide Web and we are finally embarking on fluid types of interactions that surpass traditional definitions of human-computer interfaces.

[51]The mingling of artificial and natural language is also taking on other directions. In another striking example we can see that the use of artifactual symbol systems is permitting a different sort of cross-species communication that further alters our notion of natural language discourse. Returning to one study of Koko the gorilla that Kennedy cites as "distorted by anthropomorphism," we can discern important aspects of human-primate communicative interaction that he dismisses.[52] "Project Koko" is one effort in Apple Computer's Vivarium Program headed up by Alan Kay. The problems of designing a computer-voicing mechanism for the gorilla Koko are reported in *The Art of Human-Computer Interface Design*:

> At one time, when Project Koko was still on the Stanford campus, Koko had access to a voice synthesis unit which she used to enhance her communication skills. However, when the Foundation moved to Woodside Koko lost access to this device. The loss was traumatizing – when Koko was asked what she wanted for her birthday, she signed, *Voice!* The researchers were also disappointed to lose this basic tool and have been seeking a way to reinstate this capability.[53]

Apple engineers then undertook to provide Koko with a Macintosh II to replace the voice synthesis unit. The user interface had to surmount the problems of both what might be within the dexterity of a Gorilla's hand and "a 260-pound animal running at

---

[50] Malcolm Slaney, "Lyon's Cochlear Model," Apple Computer, Inc., Technical Report #13, 1988; distributed on diskette.

[51] Jon Barwise and John Etchemendy, *The Language of First-order Logic* (Stanford, Calif.: Center for the Study of Language and Information, 1990); diskette with Tarski's World program and exercises enclosed with text.

[52] Kennedy, J. S., *The New Anthropomorphism,* p. XX.

[53] Mike Clark, et al., "Koko's Mac II: A Preliminary Report," in *The Art of Human-Computer Interface Design,* ed. Brenda Laurel (Reading, Mass.: Addison-Wesley, 1990), pp. 95-102.

about 20 miles per hour, swinging her arm with a force comparable to a 10-pound shot-put traveling at 100 miles per hour." Along with all the icon design issues, she was given a choice as to the voice she wished to be "her own." On the experimenter's side, all sorts of data recording and special programming environments were created to support the ongoing research.

Koko's use of computers is not limited to a voicing interface:

> Using SuperCard, Mike Clark developed three applications specifically to familiarize Koko with the touch screen interface. The first is four digitized animal pictures that make the animal's sound when touched. The second shows a picture for each letter of the alphabet (developed by the Gorilla Foundation) and speaks the name of the picture when it is pressed. The final application is called KokoPaint. It is a simplified color painting program which allows Koko to select from a few colors and finger paint with them.[54]

This is a merged artificial and natural language situation that goes far beyond the effectiveness of J. S. Kennedy's call for a neobehaviorism. It simply is not possible to sort out the meaningful discriminations for the network of human-computer-animal communications on the basis of an anti-anthropomorphic methodology alone. Even with the most conservative interpretations as to what Koko "really means" by her voicing, Koko is still living her life with Drs. Francine Patterson and Ron Cohn in ways that are fundamentally altered by access to these computer-assisted modes of communication.

The Mathematica™ article and Koko's voicing are but two examples of thousands that would give evidence that artificial language constructs are being incorporated as extensions to humanly embodied communications. They are to be taken as evidence that the impacts of computerized symbol systems are at least as wide as the changes to natural symbolic modes that they engender. This dissemination is not something that can be confined to the laboratory nor restricted to the social sphere of scientific communities. It may be that primary path of dissemination of these impacts will be outward from computer science communities and teams of industrial software engineers, and/or there may be patterns of influence directly between communities of users.

The determination of several modes of meaning loss detailed above makes it clear that our computerized symbolic artifacts are in need of some "broader sense of (a) 'legitimizing notion'."[55] In spite of the explosively productive payoff we receive from using computers, we cannot simply trust to the natural goodness, or even ecological closeness of their fit with us and the wider natural world. Not only do artifactual symbol systems present many dimensions of referential opacity, they also exist in a societal framework wherein no guarantee exists that all the necessary shaping factors for human viability and ecological soundness will be allowed to work upon, or even permit, successive designs. Without doubt, this sentiment is held by many, including responsible practitioners of software design. Apple Computer's unprecedented emphasis on standards and guidelines for graphical user-interface design from the first production of the Macintosh is the best case that I know of this concern put into practice. In another paper I plan to sketch the outlines of such a legitimizing notion as it might be at work in designing human-computer interactions.[56] For example, the

---

[54] *Ibid*., p. 99.

[55] Lorraine Daston, "The Moral Economy of Science," in *Constructing Knowledge in the History of Science*, ed. Arnold Thackray, *Osiris* 10, (1995): 3, n. 1.

[56] Software design is a practical discipline. My own experience with it consists primarily of managing a software development team of nine computer and content professionals over a period of roughly five years. This effort has led to a commercially published application, Biota, for modeling population dynamics. Members of the Biota design team were Jim Danbury, Ben Jones, John Kruper, Jim Lichtenstein, Eric Nelson, Jeff Schank, Bill Sterner, Joyce Weil, and William Wimsatt. For further details see B. Jones, W. Sterner, J. Schank, "Biota: An Object-oriented Tool for Modeling Complex Ecological Systems," *Mathematical and Computer Modeling*, Pergamon Press, 20: 8, October 1994, pp.

concept of “input-computation-output” is central to understand a discrete mechanical system. Yet, as we have seen it is quite inadequate to describe the human experience of communicative interactions. To fall into a design ethos limited to computational concepts would be to deny the expansions of human meaning creation made possible by digital symbols. Perhaps with considerable more research, what roles the moral economies of design play in the arts and disciplines of the designers and engineers involved can be further determined. Here I only point to where we are yet inhibited by our specialized design cultures.

The purposes of the design goals posited for a "moral economy" of artificial symbols systems arise from our need to understand and integrate the merger of natural and artificial language expressions. Four new goals in a moral economy of design were identified: (1) greater conversational expressiveness, (2) the search for theoretical completeness in human-computer interactions, (3) promoting cross-coherent symbolic worlds, and (4) enriched experiencing of symbolic agency. Through further research they may prove helpful in determining "a web of affect-saturated values that stand and function in well-defined relationship to one another" that "refers at once to the psychological and to the normative."[57]

Logic Provides Only an Objective Standard for Understanding Artifactual Symbol Systems. It Falls Far Short of Grasping the Lived Meanings Made Possible by the New Modes of Digitized Symbolization

---

31-48. See J. R. Jungck, P. Soderberg, J. N. Calley, N. S. Peterson, and J. Stewart, eds., *The BioQUEST Library*, (1994) The University of Maryland, College Park, Maryland. The *BioQUEST* Library is a compendium of computer-based tools, simulations, and textual materials to support collaborative, research-like investigations. It is distributed on CD as part of the award winning *BioQUEST* suite of biology simulation software, (EDUCOM 92/93 NCRIPTAL Awards).

[57] Daston, "The Moral Economy of Science," p. 4.

A further disciplinary or normative claim is that twentieth-century logic provides a new technical standard for the practices of manipulating symbols. It follows that the design of artifactual symbol systems in the wider sense of all the new modes of kinesthetic, visual, and auditory expressions made possible by digital technologies, as well as the incorporation of traditional verbal and diagrammatic symbolic expressions into digital technologies, must continue to work with the rigidities and limitations of formal systems. Yet this is a standard and starting point for design that we cannot permit to be reduced to preconceived notions of what a Turing machine or formal system can do. Allowing that would be an “immoral economy of design”; it would sanction a reduction of the design process to the tasks of constructing symbolic manipulations – albeit difficult, but merely instrumental and often empty – with no reference beyond those mechanical transformations to the humanly embodied meanings that the symbols take on for the people living the symbolic interactions. The reason why this is the case is that the rigorous "meaning basis" established by the discipline of modern symbolic logic has many extralogical or surplus impacts across the sciences and arts.

We can now give somewhat more stable, "fixed point" threshold for marking a human center of narrative gravity as the generative merger of user and designer viewpoints with their attendant combination of human-sensitive arts and computational arts that can mediate between what is expressible in a first-order symbol system as a symbolic artifact and the wider field of meanings that we as humans can be implicitly aware of and act from. Such a merger can also develop into a structural source of our human subjectivity as a higher order manifold of logical meaning creation.

My artistic and emotional claim is that whatever attachments we might form to our newly designed artifactual symbol systems with their vastly increased powers for algorithmic imitation, we remain closer to our own richly metaphorical natures in our interactions with other people and with animal companions than we do in human-computer interactions.

These constraints leave the imaginative productivities of computer-mediated interactions among people and with animals entirely open for development in ways about which Kant, for one, had formed no judgments.

Bibliography

Aristotle. *The Basic Works of Aristotle*, ed. Richard McKeon, New York: Random House, 1941.

Ashworth, William B., Jr. "Natural History and the Emblematic World View," in *Reappraisals of the Scientific Revolution*. Cambridge: Cambridge University Press, 1990.

Beck, Robert. "The Future of Imaging Science." Preprint for *Advances in Visual Semiotics*, 1994.

Bolter, J. David. *Turing's Man: Western Culture in the Computer Age*. Chapel Hill: The University of North Carolina Press, 1984.

Booth, Wayne C. *The Vocation of a Teacher: Rhetorical Occasions, 1967-1988* Chicago: University of Chicago Press, 1988.

Boyle, Robert. *A Free Inquiry into the Vulgarly Received Notion of Nature* London: The Works of the Honourable Robert Boyle, 1772. Reprinted in Thomas Birch ed., *Robert Boyle -The Works*. Hildesheim: Goerg Olms Verlagsbuchhandlung, 1966.

Butler, John M., ed. *Chicago University Counseling Center Discussion Papers*, (call numbers: Bf 637 C6C53), Regenstein Library, University of Chicago.

Chandler, Alfred D., Jr. *Strategy and Structure: Chapters in the History of the American Industrial Enterprise*. Cambridge: MIT Press, 1962.

Chaudhuri, Nirad C. *Scholar Extraordinary: The Life of Professor the Rt. Hon. Friedrich Max Müller, P.C.*. New York: Oxford University Press, 1974.

Darwin, Charles. *The Descent of Man, and Selection in Relation to Sex*. Princeton: Princeton University Press, (1871), 1981.

___________. *The Expression of the Emotions in Men and Animals*. Chicago: University of Chicago Press (1872), 1965.

Daston, Lorraine. "Enlightenment Calculations" *Critical Inquiry* 21 (Autumn 1994): 182-202.

____________. "The Moral Economy of Science," in *Constructing Knowledge in the History of Science*, ed. Arnold Thackray, *Osiris* 10 (1995): 3-24.

____________. "How Nature Became the Other: Anthropomorphism and Anthropocentrism in Early Modern Natural Philosophy," in *Biology as Society, Society as Biology: Metaphors*, eds. S. Maasen, E. Mendelsohn, and P. Weingart, Boston: Kluwer Academic Publishers, 1995.

Dennett, Daniel C. *Consciousness Explained*. Boston: Little, Brown and Company, 1991.

_____________. *Darwin's Dangerous Idea*, New York: Simon & Shuster, 1995.

Dreyfus, Hubert L. *What Computers* ***Still*** *Can't Do: A Critique of Artificial Reason*. Cambridge, Mass.: MIT Press, 1992.

Gellert, W. *et al*., eds. *The VNR Concise Encyclopedia of Mathematics*. New York: Van Nostrand Reinhold Company, 1977.

Gendlin, Eugene T. *Focusing-Oriented Psychotherapy*. New York: Guilford, 1996.
_____________ "Crossing and Dipping: Some Terms for Approaching the Interface between Natural Understanding and Logical Formulation," *Minds and Machines* 5 (1995): 547-560.
_____________ "Theory of Personality Change," in *Personality Change*, ed. Philip Worchel and Donn Byrne, New York: John Wiley & Sons, 1964.
_____________ *Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective* Free Press of Glencoe, 1962. (In press for republication.)

Gigerenzer, Gerd, and Goldstein, Daniel G. "Mind as Computer: The Birth of a Metaphor." *Creativity Research Journal* (In press).

Gooday, Graeme, J. N. "The Morals of Energy Metering: Constructing and Deconstructing the Precision of the Victorian Electrical Engineer's Ammeter and Voltmeter" in *The Values of Precision*, ed. M. Norton Wise, Princeton: Princeton University Press, 1995.

Harris, Roy, and Talbot, J. Taylor. *Landmarks in Linguistic Thought*. London: Routledge, 1989.

Harel, David. *Algorithmics: The Spirit of Computing*. Wokingham: Addison-Wesley Publishing Company, 1987.

Kennedy, J. S. *The New Anthropomorphism*. Cambridge: Cambridge University Press, 1992.

Laurel, Brenda, ed. *The Art of Human-Computer Interface Design*. Reading, Mass.: Addison-Wesley, 1990.

__________. *Computers as Theatre*. Reading, MA: Addison-Wesley, 1991.

D. Malament. *Notes for Intermediate Logic I & II*. University of Chicago Printing Services, 1994.

McKeon, Richard. "The Development and the Significance of the Concept of Responsibility," *Revue international de philsophie* 34 (1957, fasc. 1): 3-32.

Müller, F. Max. *Lectures on The Science of Language - delivered at the Royal Institution of Great Britain, April, May & June, 1861*. London: Longman, Green, Longman, Roberts, & Green, 1864.

__________. *Three Lectures on the Science of Language - delivered at the Oxford University Extension Meeting*. Chicago: Open Court Publishing Company, 1895.

Newell, Allen and Simon, Herbert A. "Computer Science as an Empirical Inquiry: Symbols and Search - The 1975 ACM Turing Award Lecture," in *ACM Turing Award Lectures: The First Twenty Years 1966 - 1985*. New York: ACM Press in collaboration with Addison-Wesley Publishing Company, 1987.

Parnas, David Lorge. "Software Aspects of Strategic Defense Systems," *Communications of the ACM*. 28, no. 12 (December 1985): 1326-1335.

Penrose, Roger. *The Emperor's New Mind*. Oxford: Oxford University Press, 1989.

Pinker, Steven. "Language Acquisition," in *Foundations of Cognitive Science*, ed. Michael I. Posner, Cambridge: MIT Press, 1989.

Plato, *The Collected Dialogues*. Ed. Edith Hamilton and Huntington Cairns, Princeton: Princeton University Press, 1969.

Richards, Robert J. *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*. Chicago: University of Chicago Press, 1987.

Ristau, Carolyn A., ed. *Cognitive Ethology: The Minds of Other Animals–E ssays in Honor of Donald R. Griffin*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1991.

Rogers, Carl R. *Client-Centered Therapy*. Boston: Houghton Mifflin Company, 1951.

__________. *On Becoming a Person*. Boston: Houghton Mifflin Company,1961.

Ross, Sir David. *Aristotle*, 2nd edition. London: Methuen & Co. LTD, University Paperbacks, 1949.

Simon, Herbert A. *The Sciences of the Artificial,* 3rd edition. Cambridge, Mass.: MIT Press, 1996.

Slaney, Malcolm. "Lyon's Cochlear Model, Apple Computer, Inc., Technical Report #13, 1988; distributed on diskette by Apple Computer, Inc.

Sutherland, Zena. *Children and Books*, 7th Edition. Glenview, Ill.: Scott, Foresman and Company, 1986.

Weiner, Norbert. *Cybernetics: or Control and Communication in the Animal and the Machine*, 2nd edition. Cambridge,Mass.: MIT Press, 1961.

Weizenbaum, Joseph. "ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine" in *The Communications of the ACM -(Association for Computing Machinery)*. 9, no. 1, (January, 1966): 36-45.

______________. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: W. H. Freeman and Company, 1976.

Zuboff, Shoshana. *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books, 1988.

Appendix 1 - Weizenbaum Article

(Not available on-line.)

Appendix 2 - KantPro™ Selections

The following paragraphs were generated automatically by Mark Pilgram's Kant Generator Pro™ program. The program uses a strategy of combining vocabulary and phrases selected from Kant (or anyone else) and abstracts of grammar for typical Kantian (or other author's) sentences. It then randomly selects (rather than responding to user input) words and phrases to be inserted in a sentence generating routine, churning out as much mind-numbing prose as desired. The underlying mechanism is not all that different from the one used in ELIZA, only it is not conversational.

Our sense perceptions are what first give rise to the Categories. It is obvious that, so regarded, our a posteriori knowledge exists in the objects in space and time. To avoid all misapprehension, it is necessary to explain that, so regarded, space stands in need of, however, our understanding. Since knowledge of our faculties is a priori, to avoid all misapprehension, it is necessary to explain that the things in themselves, that is to say, would be falsified; in natural theology, our faculties are just as necessary as our experience. In the study of metaphysics, I assert that the thing in itself has nothing to do with our knowledge, as any dedicated reader can clearly see. On the other hand, we can deduce that the objects in space and time can not take account of our judgments. But this need not worry us.

As is shown in the writings of Galileo, it is obvious that our judgments are the clue to the discovery of natural causes. To avoid all misapprehension, it is necessary to explain that the Categories have lying before them, in the full sense of these terms, the things in themselves; however, the Ideal excludes the possibility of the Categories. As is evident upon close examination, Hume tells us that, in particular, philosophy, so far as I know, can never furnish a true and demonstrated science, because, like philosophy, it has lying before it synthetic principles. Thus, I assert, by means of the Ideal, that the never-ending regress in the series of empirical conditions would thereby be made to contradict our faculties, by means of analytic unity. By means of analytic unity, our understanding is the clue to the discovery of our ideas, and time can not take account of the objects in space and time.

Still, Galileo tells us that the never-ending regress in the series of empirical conditions constitutes the whole content for the paralogisms of natural reason. By means of analytic unity, it is not at all certain that space may not contradict itself, but it is still possible that it may be in contradiction with pure logic. To avoid all misapprehension, it is necessary to explain that, so far as I know, transcendental logic would be falsified, yet metaphysics, still, is by its very nature contradictory. Has it ever been suggested that, by means of analytic unity, I assert that there is no relation between our a priori knowledge and the thing in itself? It is not at all certain that the never-ending regress in the series of empirical conditions, with the sole exception of the discipline of human reason, should only be used as a canon for natural causes; in natural theology, the architectonic of natural reason can not take account of the transcendental aesthetic. The objects in space and time would thereby be made to contradict, for example, our a posteriori knowledge; consequently, the noumena are by their very nature contradictory. The reader should be careful to observe that the thing in itself constitutes the whole content for, in accordance with the principles of the Antinomies, metaphysics; in the case of space, the paralogisms, as I have elsewhere shown, abstract from all content of a priori knowledge. We thus have a pure synthesis of apprehension.

The architectonic of human reason stands in need of necessity, by means of analysis. For these reasons, it remains a mystery why natural causes, certainly, stand in need to our experience. Because of our necessary ignorance of the conditions, the paralogisms are the clue to the discovery of the objects in space and time; on the other hand, the transcendental unity of apperception abstracts from all content of knowledge. The objects in space and time constitute the whole content for, in reference to ends, our experience; in natural theology, the discipline of practical reason, in view of these considerations, is a body of demonstrated doctrine, and some of it must be known a priori. The Ideal of practical reason may not contradict itself, but it is still possible that it may be in contradiction with, in so far as this expounds the practical rules of our understanding, metaphysics. The reader should be careful to observe that the architectonic of pure reason depends on the transcendental unity of apperception, as is evident upon close examination. By means of analysis, let us suppose that, in particular, our a posteriori knowledge depends on, in respect of the intelligible character, the architectonic of pure reason, yet metaphysics abstracts from all content of a priori knowledge.