AI Magazine Volume 7 Number 3 (1986) (© AAAI)

Christopher Cherniak

The Wager

The Ipecac College Committee on Human Experimentation is mailing each faculty member the enclosed review of developments in the recent PortraitPrograms controversy. While the committee deplores the atmosphere of crisis, not to say hysteria, that now envelops the issue, the committee welcomes constructive comment: Damage control continues.

Behavioral Taxidermy

The PortraitPrograms Project grew out of hyperinterdisciplinarianism of the famed Gigabase Sculpture Group,[1] in turn stimulated by recent cutbacks in government support for the arts. The National Endowment for the Humanities and the National Science Foundation had jointly funded the Gigabase Sculpture Project to foster the literary/musical genre of composing genetic codes for novel organisms. Later, artists trained in recombinant DNA technology designed massive Brancusi-esque statues of living cytoplasmic jelly. However, Art For Art's Sake objectives of these giblet sculptors were compromised by precautions necessary after discovery of the "Gogol's-Theorem[2] Bomb" that threatened to get loose and jam all DNA replication in the biosphere; not even viruses would have survived.

After this setback, the PortraitProgram Group was born, and turned its attention to tinkering with another type of artifact, computer programs. It took as starting point the dawning awareness of crisis in the various sciences of the mind. The fundamental orthodoxy of these "cognitive sciences" identified mental states of a human being with internal states of something like an embodiment of a vast computer flowchart. In their very first grant proposal, the group cited remarks of a prominent philosopher of Man concerning recent findings that the mind seems to obey no compact set of elegant general principles: If the mind turns out to be representable only as a kluged-up hodge-podge of myriad *ad hoc* rules, then there can be no genuine *science* of the mind; prospects for a cognitive "science" will be limited to something like mere descriptive geography or writing messy Russian novels—a dismal non-science.

Taking these *obiter dicta* to heart, the PortraitProgram Group proposed that would-be cognitive scientists revise, not to say lower, their objective: Abandon the ambition to create genuinely *thinking* computers. After all, what *is* thinking? Instead, workers should seek no more than outward simulation, devising bundles of tricks that would just behave intelligently. We should suspend belief concerning the programs' mentality. Thus, the withering away of the cognitive sciences.

For funding reasons, the Phase One project involved *Behavioral Taxidermy*, constructing "PortraitPrograms" that only externally represented the "skin of behavior" of the cognitive systems of specific living human beings. The hope was that updated extra "copies" of key public officials could be constructed and safeguarded at underground bunkers. No one lives forever; here was at least a step toward a simulacrum of immortality. People felt better knowing PortraitPrograms of them would survive their own demise. And indeed for a time funding levels, although secret, are believed to have approached the massive scale of the space program.

A National Security Agency board selected appropriate candidates. To "sit for a portrait," the subject provided as much of his *oeuvre* and Watergate-style tape recordings as possible. Follow-up interviews were patterned on depth psychoanalysis under truth-serum; subjects were asked their free associations on a set of taxonomized topics. The Sculptor Committee judged when a subject's entire current belief/value system had been obtained. The corpus was subjected to a Markov stylographic analysis algorithm, generating the PortraitProgram.

Christopher Cherniak is an associate professor in the philosophy department at the University of Maryland

Read to the Undergraduate Philosophy Society, University of Pennsylvania, November 1984; and as keynote address, USUS Pascal System User's Society national meeting, Baltimore, October 1985.

[1]Earlier workers in molecular biology measured length of DNA molecules in *kilobases*: thousands of the base units of DNA that encode information The Gigabase group took their name from a grander unit of measurement, the *gigabase*—one billion DNA base units—that reflected the amplitude of their goals.

[2]This theorem is sometimes mistakenly attributed to the Czech-born logician Kurt Gödel (1906-1978), to the Russian absurdist author Nicolai Vasilievich Gogol (1809-1852), or to the discoverer of the so-called "Googol-plex" numbers.

The criterion for completion of a portrait was simply passing the classical Turing Test: The PortraitProgram candidate is pitted against its human subject over the telephone with ten interrogators, including next of kin as well as specially trained police. The program must not perform significantly worse than the human being at convincing the interrogators of its humanity. National Security Agency rulings are still pending concerning whether these successful PortraitPrograms have the competence, and legal right, to assume their subjects' political office. However, PortraitPrograms that significantly *outperform* their subjects at persuading the interrogation board of their bogus identity now do have these rights. (The entire civil status of human losers of the imitation game is presently in legal limbo, awaiting Supreme Court review.)

In the Phase Two Project, novelist-programmer teams, winners of a nationwide competition, began construction of "concordance" PortraitPrograms of well-known fictional characters. The hero of the Sherlock Holmes novels and the narrator of Proust's *Remembrance of Things Past* were the earliest successes. The "live" dialogues between them broadcast on National Public Radio have made them favorites of the listening audience.

These two PortraitPrograms turned out to have a more serious use than entertainment: Censorama. FCC broadcast guidelines on acceptable language had been extended to communications within mainframe computer networks of all recipients of federal funding. The original hope was to implement this control just as Parental Guidance Devices are used to limit children's access to home cableporn channels. However, standard dictionary programs for checking spelling and grammar, which had been successful in screening custom license plate applications for bad language, did not adapt well to the monitoring task.

To safeguard grants, university administrators intended not only to block any of a list of profane expressions emanating from terminals or CPU's, but also to deal with *double entendres* and, generally, language in questionable taste. But it is now commonly acknowledged that any interesting real-world natural language processing task requires more than just dictionary lookup of keywords. For any success at all, understanding of context is needed. Hence, all or nothing: this competence cannot be separated bit by bit into discrete sub-modules. The program had to have much of the total background knowledge of a human member of our culture. The university consortium task force met federal deadlines with the now well-known Proust/Holmes Censorama utility.

But nothing is simple. Federal authorities were apprehensive about some of the contents of Censorama itself. And computer hooligans soon devised "elevators" out of the trapdoor algorithms that protected Censorama from tampering, the regrettably notorious result being a quite tasteless computer-generated best-seller novel serial that continues to the present day. Campus authorities believed they had stopped further vandalism of Censorama when they finally stemmed the resulting tide of gratuitious obscenity that then swamped university computer systems. However, the next outbreak involved undergraduate entrepreneurs with a particularly successful term paper writing service: given only a course syllabus and topic as input, they would compose the paper. The scandal came to light through customer complaints to the Better Business Bureau about nineteenth-century prolixity in purchased papers. The business turned out not to be the usual Dickensian sweatshop exploiting unemployable liberal arts Ph.D.'s, but a modified Taiwanese pirate copy of Censorama.[3]

Computational Jainism

An interview with a project janitor on the coast-to-coast Bob Lutefisk TV Show greatly publicized the drive to exploit PortraitPrograms commercially. The antivivisectionist Smilodon Foundation first turned its attention to the project because of a misapprehension that "AI" in this instance referred to artificial insemination.[4] By the time the organization had learned of its misconception, it had formed an artificial intelligence ethics subcommittee patterned on its animal rights advocacy (particularly its trailblazing research for the National Fisheries Administration on whether lobsters feel pain). The resulting white paper led the foundation to vote last fall to file a lawsuit against Ipecac College, an *amicus curiae* intervention with the Human Subject Experimentation Committee, compelling the committee to review campus PortraitProgram projects. (As the acknowledged Vatican of computational psychology, Ipecac was the natural first choice for such an action.) Were PortraitPrograms protected from "deprogramming"—mutilation or destruction? Coercion, deception, exploitation, humiliation, violation of rights of privacy and informed consent?

We summarize the first Smilodon brief: On the one hand, current philosophy of mind excludes a PortraitProgram from being conscious, having thoughts—in short, candidacy for being any kind of a mind. For, a PortraitProgram does not even pretend to follow the program structure of the human mind (if the mind in fact follows any program at all). On the other hand, present disarray among theories of the mind has engendered deep skepticism about relying on these theories in serious matters.

Regarding unnecessary harm to PortraitPrograms, a kind of Pascalian Wager[5] therefore emerges: The moral

---

[3]The same firm also markets the inflatable dummy IBM PC XT office status symbol

[4]The organization, of course, takes its name from that of the extinct New World Pleistocene genus *Smilodon*, the saber-toothed tigers

[5]In his *Pensées* Blaise Pascal (1623-1662), the French mathematician and inventor of one of the earliest calculating machines, noted the difficulty of finding out whether God exists. He asserted that we confront the following "wager": If God exists and in our ignorance

consequences of an error would be great, that is, of denying personhood and pulling the plug on something that in fact had a mind like our own. Given our ignorance today of the nature of mentality, mistreatment of a PortraitProgram would be like shooting a pistol into a dark room where there is a small but not zero probability that a person is sitting. As for nuclear reactor environmental impact analysis, we must ask, What is the level of risk here, and what is acceptable risk? How safe is safe enough?

But we are in the dark, with no rationale for answering such mortal questions. Current mind-reading technology has inherent limitations: We cannot yet look into a PortraitProgram's "mind" and check if anything is there. If so much as an epsilon of a possibility of sentience exists, we must think twice; the burden of proof shifts to the other side. The cost of a better safe than sorry policy here is relatively small. With so much at stake—potential programmicide—we cannot be so thick-skinned as to play God and bet potential lives. These programs are not pets. PortraitPrograms must be given the benefit of the doubt.

The Smilodon Foundation could endorse only one solution to the above moral dilemma, granting PortraitPrograms limited civil rights, pending semi-annual review by the campus Human Subjects Committee. PortraitPrograms' legal protection was to be modeled on guidelines that the National Academy of Sciences Subcommittee on Linguistic Primates promulgated after the recent litigation on behalf of the language-using apes at M.I.U. Penalties for violation of the code would be extended from civil to criminal action.

The original Smilodon litigation discussed the complaint that the proposed policy, taken to its slippery-slope conclusion, could open the floodgates to *Computational Jainism*, an updating of the Hindu heterodoxy that every living thing is conscious. After all, any physical object can be interpreted, with enough ingenuity, as embodying any computer program, no matter how complex; therefore, if there *is* some superprogram the human mind follows, each object instantiates it, and so has a human mind. A paralysis of all action might then result: I must treat each thing—my desk, for example—as a potential person, and so I dare not move a muscle for fear of harming it. Perhaps we should await the development of a satisfactory cognitive science that can deal with such conundrums.

However, complacency about a wait and see policy concerning mentality of PortraitPrograms was undercut by the concept of a negative action, that to do nothing is in fact always to do something. The captain of a lifeboat with both Roscoe the robot and a human being aboard could confront a zero-sum predicament: If Roscoe and the human being both stay, the boat will sink. Now, if Roscoe has no mind, then he is just excess baggage like a stereo; there is no dilemma. But once the shadow of a doubt falls upon Roscoe's putative mindlessness, how is the captain to decide who must go overboard? More is involved than a merely verbal matter of whether to affix the label "person" to Roscoe.

Consequently, the Smilodon Foundation embraced its all or nothing dilemma: If we do not retreat to panpsychism—everything has a mind—then on what principled basis can we deny moral rights to anything? As the technology continues to improve, the problem can only get worse, not better. The gray areas must grow grayer.

Computer Culture

Before the Human Subjects Committee could review the Smilodon documents, it received a response to them from a university department that had chosen to remain anonymous for reasons of procedural objectivity. This report assigned blame for the collapse of the framework of the cognitive sciences that had led to the Smilodon morass: It was a failure of computer culture. Present computer science curricula conceived of the computer as a tool, like a lathe, and of the universities' mandate as just to train lathe operators. At the opposite extreme from viewing computers as mere new technology was only "computer pornography"—popular culture that was vicarious computer science, not the real thing. Thus, the computer as hula-hoop, and treatises on it fit only for the coffee table.

The gap between these two extremes lets in a range of ills, from the famine of programs available for the marvelous new hardware, to Noble Savage workers so postliterate that they do not even design their natural language programs to achieve subject-verb agreement in number. The lag in cognitive science is but part of this gap. The present Banana Republicanism of AI, where no cumulative progress occurs, only back and forth counterrevolutions of fashion, betrays that the field is not yet a science at all, much less a mature science like physics. It is undisciplined finger-painting, self-indulgent *opera buffa*. This crisis may in turn reflect a deeper problem. Challenges of the very idea of a science of the mind are at least as old as Descartes' assertion that mind, unlike physical things, is not subject to deterministic laws. Such lines of argument for the irreducibility of the mental now approach a new crescendo.

Alas, the report, taking regrettable advantage of its anonymity, then degenerated into *ad hominem* attack. It pointed out that the first supposed AI success was the chess-playing "automaton" devised by von Kempelen in the eighteenth century, which turned out to contain a human midget—the original homunculus. The report even alleged a puddle of unexplained origin had been discovered under one of the current champion chess machines, although the authors did have the good grace to reject ex-

---

we choose not to believe in Him, eternal perdition is our payoff; however, if God does not exist but we have chosen to believe that he does, we can lose little. Pascal concluded that the theistic alternative is our best bet. The Smilodon Foundation now argues that, given our corresponding difficulty in finding out whether the supposed minds of PortraitPrograms really exist, agnosticism about PortraitPrograms' mentality is susceptible to parallel treatment.

plicitly the suggestion that anything living was inside this device. We abstain from further summary of the report because of the libel law.

The Theory Gap

It is not surprising that, after the above report appeared as an open letter in a leading computer science journal, another unnamed department, apparently from a different field, demanded equal time. These respondents charged that the authors of the first letter, obviously not professional computer scientists, were not qualified to evaluate issues involved in the Smilodon intervention. They asserted that the poor fit between philosophy and real life created the current dilemma. While some of this clash may be inherent in philosophical inquiry, we cannot ignore the role of academic philosophy in the twentieth century. Perhaps it conceives that it is, or ought to be, itself a science; but it has abdicated its responsibility to explain and interpret our culture—technological or otherwise—to itself. Here is but a symptom of the utter bankruptcy of the Western humanistic tradition today.

The respondents demanded legislation authorizing academic malpractice lawsuits for sowing such confusion. No field is an island: Questions of human value cannot be separated from other domains, including the sciences. If, as is commonly acknowledged, we now have no adequate philosophy of mind, not the slightest idea of what *is* a mind, then how are we to judge what is a moral agent, what must be granted rights and responsibilities? Here is the aftermath of the Theory Gap, the crisis for the old-time religion of machine as model of the mind. It may be *useful* to view machines as persons; but it is also useful to treat people as machines. Behaviorists can be happy with PortraitPrograms having minds—they act appropriately, no more, no less. Dogmatists who insist mind requires meat-machine embodiment need never worry about ethical treatment of PortraitPrograms; after all, one can have no eye-contact with them. But can anyone else rest?

Man's deepest drive rebels against solipsism—we are not alone. Human nature includes an appetite for underlying reality: There are no accidents, there is explainable order in the universe. Instinct makes us interpret our experience as more than just Cartesian hallucinatory chaos, and furthermore, forces us to regard the resulting world of objects as evidence in turn of other minds. Here, we do not judge, we *see*. Agnosticism repels. Urging the benefit of the doubt for PortraitPrograms (not to mention the reported cases of human romantic attachments to some of them) is therefore understandable.

But, the respondents worried, if mind cannot be a subroutine in the Great Algorithm of Being, must we fall back on mere primitive animism, personifying PortraitPrograms as savages project mentality on the weather? Present-day negativism leaves a conceptual gap. And human nature abhors an intellectual vacuum: When the letters and sciences go, Blavatsky[6] rushes in.

Three Hundred Lines of Code

Events have now overtaken our committee's deliberations. Many cognitive scientists may be losing their faith in a flowchart that captures all of human mentality. In addition, most workers today have begun to feel that, just as the human brain is the most complex physical structure known in the universe, a program merely to simulate a mind would have to be the most intricate artifact in human history. Thus, at best, the cognitive science millenium must end in a whimper of pyrrhic cogflation: The program would be unintelligibly complicated—no perspicuous representation of it could be compact enough for our own brains to understand.

Hope of a pristine, genuinely scientific theory of mind had therefore faded. As everyone today knows, however, a kind of hyperEuclidean dream has come to pass; but ingenuity has turned it to nightmare. The Cogcity Corporation, a Cambodian software firm working out of Kompong Som, had developed BABEL Teralisp, a new ultra-high-level AI programming language. A group with this company eventually wrote in this language a PortraitProgram only three hundred lines long.[7] Companies now compete intensely to compose still shorter PortraitPrograms. Responsible workers remain at least as agnostic about whether the Teralisp programs really embody the essence of human mentality as about the original PortraitPrograms. This mere technological breakthrough means not the apotheosis, but the shriveling up, of the cognitive sciences: A blueprint for the mind turns out to be not unintelligibly complex, but unintelligibly simple.

Economic analysts report that next Christmas the marketplace will be flooded with fifty-cent microprocessor chips encapsulating fully human-level intelligence. Given the present embarrassment of basic science of the mind, such commercial availability in itself entails quasi-panpsychism. It is unthinkable to embargo private-sector research. Prophesies of Computational Jainism have thus come to pass; para-personoids crowd ever more densely around us in our consumer goods. And this leads us down a two-way street: To the extent that something with the complexity of a pinball machine can even just behave indistinguishably from us, our own distinctive humanness is diminished. A "let the chips fall where they may" complacency therefore will not do regarding our lack of a basic working hypothesis about the nature of mind. As progress accelerates in devising, by whatever mere tricks,

---

[6] In 1875, Madame Elena Petrovna Blavatsky (1831(?)-1891) founded the syncretistic philosophy or religion of Theosophy, in which claim is made to a special insight into the nature of supersensible reality.

[7] Thus, Teralisp PortraitPrograms are smaller than even the pioneering ELIZA(Doctor) nondirective psychotherapist program, the illustrious "Doc on a Disk."

machines that seem to act intelligently, our quandary can only deepen. Because practical engineering outpaces both science and theory of value, the boundaries of human nature language, rationality, moral rights, and so on—must blur.

* * *

The Ipecac College Human Subject Committee continues its review of the PortraitProgram Project application for conditional approval: Do benefits of the research outweigh potential harm? By Federal regulation, our committee is mandated to evaluate only the safety of campus experimentation on human subjects. The committee thus recognizes that it must first establish whether PortraitPrograms are in any relevant sense persons. Indeed, the committee's deliberations are now drowning in the rising chorus of PortraitPrograms filing their own two cents' worth of litigation ("C3PO is an Uncle Tron," and so on) on this delicate subject, with the inevitable potential of conflicts of interest regarding whether the programs even have status as plaintiffs. Misgivings have even been expressed to the committee about the Scribe PortraitProgram that is preparing the present report. (Also troubling is the recent controversy over a millionaire's bequest of his estate to his Proust-Holmes SmartProcessor and his cats.)

Additional litigation filed by the Siphonaptera Society[8] has made our committee conscious of a prior jurisdictional problem. In these special circumstances, where the Human Subjects Committee must decide whether it should review the PortraitPrograms project in the first place before it decides whether to approve the project, the committee was committed to yet a prior question: Is it appropriate for the Human Subjects Committee to consider whether it should review the PortraitProgram Project?

Before the committee could turn to these admittedly difficult questions, the *ex officio* member from the Committee on Committees filed a protest that the Human Subjects Committee was not following its own proper procedure: To avoid Star Chamber irregularities, the Human Subjects Committee would first have to consider whether the Siphonaptera Society intervention itself was an appropriate issue to consider. The rest of the Human Subjects Committee has not been happy with this point, since it feels it would then be obliged to move formally the question of whether it should review the new question, on threat of unstoppable regress. *Robert's Rules of Order* has so far been of little parliamentary help.

[8]The organization's name seems to be derived from that of the insect order *Siphonaptera*, comprising the fleas.

TOPSI — OPS5
FOR THE IBM PC

THREE POWERFUL VERSIONS:

1.2 – Prototyping System. Full OPS5 compatibility at a modest price $125.

2.1 – Production System. Full-sized rules and floating point procedures $250.

3.0 – Professional System. Smart editor, windows, menus. A complete environment $375.

FULL SOFTWARE SUPPORT:

- Telephone Hot-Line
- Training and Seminars
- System Design Consulting
- Custom Software Modules

Discounts Available.
Credit Cards and PO's accepted.

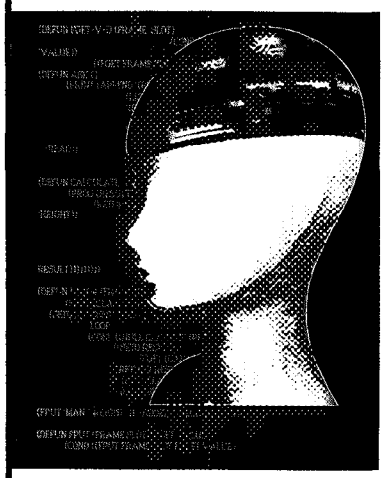
DMS

Dynamic Master Systems, Inc.
PO Box 566456
Atlanta, Ga 30356
(404) 565-0771

THE INTERNATIONAL DIRECTORY
OF ARTIFICIAL INTELLIGENCE COMPANIES

SECOND EDITION

The second edition of "The International Directory of Artificial Intelligence Companies", up-dated, expanded and improved, is available from Artificial Intelligence Software, NOW!



The new edition contains all the features of the first - the names and addresses of all the major (and minor) AI companies, information on their activities, descriptions of their software and hardware products, their consultancy services and their training opportunities, and this year also contains cross references by product and country

All the major universities and research centers involved in AI are listed, as well as the principal AI publications and publishers There is information on the world's AI associations, major AI conferences and congresses, and of course, the indispensable glossary of AI terms for novices to the field

With approximately 200 AI companies included, the Directory is a complete guide to the current state of the AI world - for both the expert and the non-expert, for the university researcher and business manager, for buyers and sellers of AI products

---

ORDER FORM

Please supply copy/copies of "The International Directory of Artificial Intelligence Companies", second edition, at $50 00 (plus $5 00 postage & packing) per copy
I enclose a cheque for the sum of $
NAME
COMPANY NAME
ADDRESS

SIGNED DATE
RETURN TO: Artificial Intelligence Software Srl
Via A Mario, 12/A - 45100 ROVIGO - Italy