

Religious Studies Independent Project:
Artificial Intelligence and Human Destiny

An Independent Project inquiring into the theological and ethical implications of current research endeavors in Computer Science, their prospects for success, and the meaning such success would have for the future of mankind.

I am, of course, aware that the project described above is sufficiently broad for a lifetime of research; in fact, I fully expect to devote many years to it myself. Thus the problem is one of limitation, of defining a sufficiently small area as the topic of a semester's independent work and an extensive final paper. For the current project, then, I shall ignore as much as possible the technical questions of ultimate computer capabilities, the psychological questions regarding the motivations of the men who strive so mightily to create a machine that surpasses its makers, and the theological questions inherent in any such human attempt to play God. Instead, I shall concentrate on two key religious questions: First, what would be necessary in a computer for it to be definitively said to have duplicated or surpassed Man? Second, assuming that such computers may ultimately surpass humanity, what, if anything, should be done to pass on to them human morality and the highest values of mankind?

The first question is really a modern phrasing of the ancient question, "What is Man?" It is a question many

religious thinkers have dealt with, and my task will be largely to glean from their writings a set of criteria by which Man may be set apart and which might be set as standards for "intelligent" machines. The second question is more difficult; indeed it may be absurd to even contemplate an attempt to mold the morality of an entity that will be, at a minimum, our equal. Yet as parents inevitably affect the morality of their children, so too we can not be merely a neutral influence on our own creations. Here I will consult the many religious speculations on the future course of evolution and perhaps some psychological writings on the passing down of ethical values.

A partial bibliography follows. Inevitably I have already read many of the books that are on my current list. I am sure that there are many books of equal or greater relevance to my subject that I will discover in the coming months.

Bibliography

Aurobindo, The Future Evolution of Man.

Barrett, Irrational Man.

Bloch, The Golem.

Buber, The Knowledge of Man and I and Thou.

Dreyfuss, What Computers Can't Do: A Critique of Artificial Reason.

Heschel, Who is Man.

Laszlo, The Systems View of the World.

Martin & Norman, The Computerized Society.

Raphael, The Thinking Computer: Mind inside Matter.

Skinner, Beyond Freedom and Dignity.

Tagore, The Religion of Man.

Weizenbaum, Computer Power and Human Reason.

A Proposal
for the 1979-80 Thomas J. Watson Foundation
Fellowship Program
"Artificial Intelligence and Human Destiny:
a cross-cultural perspective"
by Nathaniel S. Borenstein

I. Purpose

"When we start thinking of men as automaton, clicking their respective ways through the process of life with mechanical exactness, that day we lose our own identity and become automaton ourselves."

-- Thomas J. Watson, 1926. (1)

No one really knows what the ultimate limits of machine intelligence may be, or even if such limits exist. Among computer scientists there is widespread belief that artificial minds will in many ways rival Man's in the not-distant future. Yet most laymen have no such expectations, and would doubtless find the existence of intelligent machines deeply troublesome. If the scientists are correct, this trauma is inevitable. And even if such machines are not lurking around the corner, mankind will soon have to face a world in which many former human functions become swiftly automated. All of this poses enormous philosophical and existential problems that have not, by and large, been given serious philosophical consideration. Such discussions as

have taken place have been mostly serious but brief inquiries by concerned computer scientists in their spare time, or half-serious discussions by technically unqualified philosophers and philosophy students who find it difficult to believe that what computer scientists take for granted can ever really come to pass. Yet these are weighty and even vital matters; indeed the idea that machines may surpass human intelligence challenges the most crucial philosophical and religious assumptions about the destiny of mankind. Elevating the level of their discussion, indeed the very establishment of a philosophically rigorous and technically informed enquiry into these issues, has become my major life's goal.

II. Topic

"What can one do about which one can feel optimistic? Very little.... Whom do I know who seems optimistic? There are not many.... Computer people are because the gallop of their technology will soon put the whole of man's social life into their hands. And more: they will be able to encapsulate the seed of intelligence in their chips and release it from the inching of evolution. It took two hundred million years to get us here: in maybe a thousand more man will be done for, and the vital spark he nourishes so dimly will be on its way alone, unencumbered by the flesh."

-- Peter Laurie, 1979. (2)

It is against the background of these larger concerns that I have come to formulate a more modest goal for a year's research under the auspices of the Watson Foundation. I propose to investigate the expectations, hopes, and fears that computer scientists themselves hold with regard to the future of artificial intelligence. Clearly computer scientists are currently in a better position to evaluate that future than

most laymen. Yet I have found that, perhaps because of their general lack of philosophical training, American computer scientists tend to imbue their speculations on the future with ethnocentric images of human nature and human needs. To hear them talk, the computer often seems the embodiment of the noblest Western values -- but would an Indian computer programmer view it the same way? My goal is to seek out the common ground, to find what is expected by computer scientists in many cultures, and thus to isolate in the purest form possible the basic human response to the computer. At the end of the year, I hope to have a broader and deeper philosophical base from which to plunge into a graduate study of computer science -- a study which, in turn, I shall undertake in order to obtain a firm technical competence before venturing into serious philosophical endeavors. Before I can inquire into what computers will mean for mankind, I must learn well what they can actually do. But before I immerse myself in the study of the machine's capacities, I wish to know the biases, hopes, and fears of the men who build and work with them. It is for this step that I seek the Watson Foundation's support.

III. Plan (Methodology)

"Intelligence is a meaningless concept in and of itself. It requires a frame of reference, a specification of a domain of thought and action, in order to make it reasonable."

-- Joseph Weizenbaum, 1976. (3)

My plan is a simple one. I will go to the largest urban centers in various countries and seek out computer professionals,

especially programmers, the people who know the machines most intimately. Contacting these people should not be difficult. Even without prior preparation, one could easily find major computer installations by going to the banks, universities, and large businesses and asking to talk with the director of computing operations. Such installations are found even in the poorest of countries. The process may also be expedited substantially by the fact that I am seeking help from IBM, Xerox, DEC, and other multinational corporations that use computers extensively. It is my hope that these companies will tell me where I can go in various countries to interview their personnel, and perhaps will smooth the way for me by telling their local computer centers to expect me or arranging for me to meet specific people. In any event, once I find the people to interviews, I anticipate no further difficulties. I have some interview experience from my days as a highschool journalist, and besides, once found, computer programmers generally turn out to be talkative and friendly, especially when you want to talk about their machines. Of course, many companies may employ American programmers, but in all countries the number of native programmers is growing, and it is them that I shall seek out. It is a safe assumption that they will speak English, the language of computers the world over, so I foresee no insurmountable language barriers. Additionally, it may prove useful that I speak Hebrew and Spanish, as well as a little German and Italian.

Once I am actually communicating with programmers from other cultures, I will, as I said, seek to determine their view

of the meaning of computers for human destiny. In these conversations I will seek to be as flexible as possible, pursuing the topics of greatest concern to the programmers themselves. Still, I will not hesitate to introduce the topics that concern me most; I can always draw on the list of questions that I have been compiling in my notebooks with regard to computers and human destiny. (A representative sample, gleaned from those notebooks, is attached to this proposal.) I will record these conversations on tapes for future reference, and will try to classify and compare responses once I have accumulated a wide cross-cultural sample. A satisfactory result of the year's endeavor, in my view, would be a list of fears, hopes, and "scientific expectations" that transcend cultural boundaries and are widely held among computer scientists. I would then feel better prepared to learn the actual facts and theories of the academic world of computer science.

IV. Qualifications and Preparations

"A talent for manipulating symbols tempts its possessors into habitual symbol manipulation, and habitual symbol manipulation is an obstacle in the way of concrete experiencing and the reception of gratuitous graces."

-- Aldous Huxley, 1962. (4)

My interest in computers dates to my discovery, in tenth grade, that the IBM computer to which my school had a remote terminal link-up seemed to know more about practical mathematical reasoning than my high school mathematics teachers. While this did not, to my mind, make the machine intelligent, it certainly whetted my curiosity and gave me something creative to do in high school. By the time I reached Grinnell College,

computing had become a hobby; it quickly became a source of income as well, and I have been programming for the College since my freshman year, including a system to automate reserve book circulation in the college library. Still, I retained an almost complete aversion to devoting my life to punching out code for a machine, and it was probably this aversion and the apparent meaninglessness of such an existence that led me to ask the questions and take the courses that made me a Religious Studies major. My dominant concern was, and is, the meaning of human existence on this planet. But reading Sri Aurobindo's speculations on the future evolution of man turned my mind inevitably to the computer -- could it be the path through which Mind is evolving beyond Man? I was continually frustrated by the way my fellow Religious Studies students would dismiss such speculation as nonsense without the vaguest inkling of the powers and limitations of computers. But I was equally frustrated by the attitudes of my fellow Math students -- I declared a second major in Mathematics in my sophomore year -- who generally considered computer intelligence not only possible but an unservedly good thing to be striven towards unhesitatingly, and without wasting valuable time in philosophizing. Most disturbing to me was the almost total divergence in expectations; many people, inevitably, will be extremely surprised by whatever computers do or do not do.

These growing preoccupations have found more and more of a place in my academic work. In my Junior year, I received extra credit for "course intensification" in a course on "Problems in Modern Jewish Thought". The extra credit was for a

39 page paper entitled "Judaism and the Ethics of Artificial Intelligence". In my Senior year I am taking a broader view with a four credit independent study, "The Religious Implications of Artificial Intelligence". Finally, I expect to graduate not only with a double major in Mathematics and Religious Studies but also with an Interdisciplinary Concentration in Computer Science.

Such, then, are my qualifications to undertake this project. In addition, there are certain preparations I am making and will make before the actual year abroad begins. As I said before, I am compiling an ongoing list of interview questions, a sample from which is attached, and I am writing to various multinational corporations for assistance. In addition, I am writing to various embassies for advice and assistance in undertaking this research in their countries. Finally, I hope to put together a file of clippings regarding artificial intelligence that I can use to clarify questions of fact in the course of my interviews. In general, I think that the best preparation I can undertake is to broaden and deepen my knowledge as much as possible. I am discussing computer-related issues with anyone who will listen.

V. Tentative Itinerary

I know exactly where I would like to go, but the nature of my project is such that there are several valid alternatives for each country. I am thus writing to the embassies of several "alternate" countries, and am prepared to alter my

itinerary somewhat if one of these countries offers me an unusual amount of guidance or assistance, or if problems develop with my visiting one of the initially selected countries. Most important to me is that I visit countries of widely differing cultural backgrounds. I hope to visit six countries.

1. India (Alternates: Sri Lanka, Nepal)
2. China (Alternates: Taiwan, Japan, South Korea)
3. Venezuela (Alternates: Columbia, Chile, Costa Rica)
4. Roumania (Alternates: Yugoslavia, Poland)
5. Egypt (Alternates: Sudan, Turkey)
6. Kenya (Alternates: Saire, Nigeria, Uganda, Zambia)

VI. Sample Questions for Interviews

The following are among the many computer-oriented questions in my notebooks:

- To what extent can computers ever successfully imitate human intelligence?
- Are there human capabilities that can never be replaced by machines?
- How will people cope with limitless leisure time?
- Are computers the next evolutionary step after Man?
- Can a computer learn to choose freely?
- Are there capabilities which can be given to computers but ought not to be?
- Who is ethically responsible for the errors of a complex program in an unusual and unforeseen situation?
- What are the major limitations on computer thought?
- What are the major limitations on human thought?
- Can we pass on to machines our highest moral and ethical ideals?
- Can human beings cope with being second-best? What will men do if and when machines begin to surpass them in nearly every way? Does mankind have a destiny beyond the production of computers?
- What good will computers bring to mankind on earth? What evil?
- Towards what do you, as a computer scientist, feel you are working?
- How can a computer separate the relevant from the irrelevant?

VII. References

1. Dr. Henry Walker, Professor of Mathematics
2. Dr. Harold Kasimow, Professor of Religion
3. Dr. Thomas Moberg, Professor of Mathematics
4. Mr. Steven Pandolfo, Librarian

VIII. Notes

1. Quoted in THINK: A Biography of the Watsons and IBM, by William Rodgers, New American Library, New York, 1969.
2. Peter Laurie, "Pig Ignorant: Who is Optimistic?" in Coevolution Quarterly, Number 22, Summer 1979.
3. Joseph Weizenbaum, Computer Power and Human Reason, San Francisco, W.H. Freeman and Co., 1976.
4. Aldous Huxley, Island, Harper and Row, New York, 1962.