Stanford Univ
Stan-CS-76-577

S 252

Review of Joseph Weizenbaum's Computer Power and Human Reason

by
Joshua Lederberg

Professor of Genetics
Stanford University

The following review of J. Weizenbaum's book was solicited and accepted by the N.Y.Times Book Review 1 March 1976 but to the best of my knowledge has not appeared in print. J. Lederberg

"Computer power and human reason" is a mosaic of well-reasoned analysis and passionate pleading on the nature of computers and of man, and about the place that computers (read "technology" if you wish) should have in human affairs. Prof. Weizenbaum is particularly exercised about the claims for and prospects of AI, "artificial intelligence", the efforts to emulate and bolster human reasoning processes as programs in computers. A well known computer scientist at M.I.T, who has made significant contributions to AI, he writes that he was moved to write this tract when people responded to one of his programs as if it were an empathic companion .This over-estimation of, and over-dependence upon computers,he believes to be both symptom and cause of global predicaments with more horrors to come.

Weizenbaum may still be too much a technocrat: witness his oversimplifications of social movements as the immediate fruits of technical innovations. (There is more to the history of the internal migration and urbanization of American blacks than the introduction of the mechanical cotton-picker in the 1950's.)But to dwell on these would do too little justice to the other fundamental issues that Weizenbaum raises.Indeed he might be the first to deplore his own vestigial technocratic biases, when they are inconsistent with his fundamental ethical philosophy. Nor should one harp on his ad hominem attacks on some of his colleagues, his bludgeoning them with selected quotations from writings of 20 years ago,which I suspect he will view as a lapse more from enthusiasm than from malicious intention.

Most readers will follow the author's advice to skip over the early chapters which detail the fundamental logic of the computer -- these would make another, peerless book for explaining Turing's work on the fundamental logic of computing machines to the lay reader. The basic philosophical and policy issues do not need this detail, and are best scrutinized by reading the book back-to-front: few readers with the interest and general intellectual grounding to digest this work

critically, will need to be reintroduced to the fundamentals. Others will be attracted by pages of lyrical anti-technology slogans, which the author's technical reputation will make the more persuasive. Since the author categorically rejects "instrumental reason" in its application to human affairs, it is difficult to engage him in a discussion of his particular policy concerns.

Weizenbaum makes a conscientious effort to distinguish his assertions of faith from the scientific consensus; but the non-specialist reader will still have to look closely to be sure. Perhaps in fields like the physiology of the right versus left brain he has already persuaded himself that contemporary speculations are proven realities about the location of human rational functions. But others should be cautioned that we still know even less about the organization of human intellect than Weizenbaum stipulates.

During the early adolescence of computer science in the early 50's, many workers made extravagant prophecies about the ease with which the new machines would be programmed to match human problem-solving behavior -"within the visible future", we were told, machines would conduct mechanical translations of high quality (especially from Russian into English). They would play chess to the disadvantage of the masters, and they might then be ready to take over many of the higher-level functions of management in industry, and of command and control in the military. Within a few years, the power of machines to manipulate bits of information had been enhanced a million-fold: what more could one ask as the basis for these new powers?. It is no surprise, and by now no news, that these prophecies were simply wrong; and the wiser among us should have learned not to make technological forecasts where we simply had new tools, but no real insight into the structure of the tasks they were to address. Surely, as Weizenbaum insists, there are few things less well understood than human creative imagination. His own prophecy is that this will NEVER be emulated to any significant measure by computing machines. This hypothesis is beyond the range of scientific criticism, short of tangible advances too much to hope for right away; but his arguments are mainly repetitious assertions of his personal faith.

No, there is one more persuasive kernel: namely that the world-knowledge which underlies human understanding (compassion and judgment) needs the life-long experience of having been human -- in a word, of having shared love .It is unlikely and undesirable that machines be offered that privilege; then many realms will be uniquely human. Indeed we must make equally sure that the fellow-creatures to whom we confide our trust for ethical and esthetic leadership justify this on the same grounds. The abrogation of human responsibility for moral decision whether it be out of lazy delegation to machines, or superstitious deference to super-human abstractions can indeed once again ignite the holocaust.

Weizenbaum's pleading overreaches this sufficient argument to an out-and-out obscurantism about the fundamental non-comprehensibility of the human brain, which adds little to the debates between

vitalists and mechanists of the last two centuries. It is a sterile debate; and scientists can contribute more by trying to find what can be learned about our own nature, and putting it to human good, than arguing what may or may not be ultimately knowable. As with his concern about the bounds of AI, the mischief of such criticism is that it may disparage the work of investigators with more concrete, modest and achievable goals. The view that the core of the cell's reproductive capability was unknowable in chemical terms bears much of the onus for long delays in our understanding of the structure of DNA. After the fact, this proved to be remarkably simple.

While we should not offer love to the machine, there is much to be said for permitting it to evolve, that is to nurture the growth of more and more complex programs. These are initiated by human intelligence, but grow from the dynamics built into the starting program itself. It is hard to see how some of the more complex problems to be addressed can be solved by programs that are explicitly written in detail by human authors. Then I agree with Weizenbaum that we can longer claim to have a full-fledged explanation of a phenomenon, merely through having generated a model for it. We may even have substantial power to solve problems without necessarily "understanding" them. (Unfortunately, this criticism does little to help us recognize true understanding by any objective criterion). Furthermore, we should not trust such complex programs merely because we believe we were sufficiently intelligent in our original design plans. Instead, the program will have become another experiment, to be validated only by experience. Much the same ought to be said for other areas of human aspiration, like politics.

Weizenbaum is particularly critical of the use of the computer in the role of psychotherapy -- doubtless in consternation that the machine's patrons believed they were talking to a sympathetic, understanding 'person'. This criticism raises a number of issues that deserve more analytical attention: 1) Is it true that the patrons were confused, or do many of them find some service in the 'dialogue' well knowing that they are at best talking to themselves? and 2) Can a therapeutic utility by this modality be empirically validated and economically justified? At the moment our answers to these questions are speculative and anecdotal, and I would not substitute my own critical skepticism about this approach for a conclusive dismissal of it. Weizenbaum's criticism indeed may misapprehend the role of psychotherapy as a source of self-insight -- where the patient himself must do most of the work at achieving human understanding -- and machines may well be expected to play some useful role in this process (no differently than, say, the reading of a book -- and in a similar analogy to computer-assisted-instruction in other domains.)

Throughout his book, Weizenbaum oscillates between a disparagement of the potential and actual accomplishments of AI, dismay at what he sees as excessive faith and dependence on this technology, and concern for some potential abuses of its development, should it be realized. That policy-makers, the public, and computer scientists alike should take a more critical and pragmatic view of the field than the zealots of 20 years ago may be granted; many well-informed people within the field clearly do, without having reacted as strongly as Weizenbaum.

The abuses might be either ideological or technological. If human intelligence were more successfully mirrored in the machine, will that not justify treating human beings as if they were MERE machines? His position on this issue is colored by the experience of Nazi Germany; but the argument is confused. The most savage tyrannies that I can find in history, including Nazism, had no doubt about a unique elan-vital -- just that one folk or credo had more than an equal share. People who are philosophically concerned about the mechanistic basis of life are also overawed by its complexity, and too concerned about learning more about it to occupy themselves with holy wars. They are the least likely to be sacrificing either people or machines on the grounds of ideological conviction.

The historical record is less reassuring about the augmentation of power in the hands of irrational man: we can still argue about the case against Prometheus,Gutenberg, Galileo,or Faraday -- not to mention Oppenheimer -- but by that very token, I do not share Weizenbaum's confidence in deciding which innovations are dangerous. He points to very real concerns about machines that could interpret speech (while denying their feasibility). Yes-- they might make large scale wire-tapping irresistible, and perhaps undo the virtues of the telephone as a medium of private communication. They might also relieve millions of office-persons from the mindless tasks of transcribing the words of others, and free them for more creative responsibilities. Both of these contingencies lay heavy burdens on the adaptability of our social institutions, and it is important that we be alerted to them.

Weizenbaum does point to projects in mathematics and chemistry where computers have shown their potential for assisting human scientists in solving problems. He correctly points out that these successes are based on the existence of "strong theories" about their subject matter. We can agree that "common sense" is the human competence hardest to copy in a machine, and that the most constructive advances should come from the wisest division of labor in a synergism of man and his machines.Computers will not give us magical answers to the problems that we, or they, create: with sweat and insight we may be able to develop them as ever more effective tools to serve human needs.