COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT TO CALCULATION

Joseph Weizenbaum

THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

W. H. FREEMAN AND COMPANY
New York San Francisco

CONTENTS

INTRODUCTION

In 1935, Michael Polanyi, then holder of the Chair of Physical Chemistry at the Victoria University of Manchester, England, was suddenly shocked into a confrontation with philosophical questions that have ever since dominated his life. The shock was administered by Nicolai Bukharin, one of the leading theoreticians of the Russian Communist party, who told Polanyi that "under socialism the conception of science pursued for its own sake would disappear, for the interests of scientists would spontaneously turn to the problems of the current Five Year Plan." Polanyi sensed then that "the scientific outlook appeared to have produced a mechanical conception of man and history in which there was no place for science itself." And further that "this conception denied altogether any intrinsic power to thought and thus denied any grounds for claiming freedom of thought."[1]

I don't know how much time Polanyi thought he would devote to developing an argument for a contrary concept of man and history. His very shock testifies to the fact that he was in profound disagreement with Bukharin, therefore that he already conceived of man differently, even if he could not then give explicit form to his concept. It may be that he determined to write a counterargument to Bukharin's position, drawing only on his own experience as a scientist, and to have done with it in short order. As it turned out, however, the confrontation with philosophy triggered by Bukharin's revelation was to demand Polanyi's entire attention from then to the present day.

I recite this bit of history for two reasons. The first is to illustrate that ideas which seem at first glance to be obvious and simple, and which ought therefore to be universally credible once they have been articulated, are sometimes buoys marking out stormy channels in deep intellectual seas. That science is creative, that the creative act in science is equivalent to the creative act in art, that creation springs only from autonomous individuals, is such a simple and, one might think, obvious idea. Yet Polanyi has, as have many others, spent nearly a lifetime exploring the ground in which it is anchored and the turbulent sea of implications which surrounds it.

The second reason I recite this history is that I feel myself to be reliving part of it. My own shock was administered not by any important political figure espousing his philosophy of science, but by some people who insisted on misinterpreting a piece of work I had done. I write this without bitterness and certainly not in a defensive mood. Indeed, the interpretations I have in mind tended, if anything, to overrate what little I had accomplished and certainly its importance. No, I recall that piece of work now only because it seems to me to provide the most parsimonious way of identifying the issues I mean to discuss.

The work was done in the period 1964–1966, and was reported in the computer-science literature in January 1966 and August 1967. To summarize it briefly, I composed a computer program with which one could "converse" in English. The human conversationalist partner would type his portion of the conversation on a

typewriter connected to a computer, and the computer, under control of my program, would analyze the message that had so been transmitted to it, compose a response to it in English, and cause the response to be typed on the computer's typewriter.

I chose the name ELIZA for the language analysis program because, like the Eliza of Pygmalion fame, it could be taught to "speak" increasingly well. Because conversations must be about something, that is, because they must take place within some context, the program was constructed in a two-tier arrangement, the first tier consisting of the language analyzer and the second of a script. The script is a set of rules rather like those that might be given to an actor who is to use them to improvise around a certain theme. Thus ELIZA could be given a script to enable it to maintain a conversation about cooking eggs or about managing a bank checking account, and so on. Each specific script thus enabled ELIZA to play a specific conversational role.

For my first experiment, I gave ELIZA a script designed to permit it to play (I should really say parody) the role of a Rogerian psychotherapist engaged in an initial interview with a patient. The Rogerian psychotherapist is relatively easy to imitate because much of his technique consists of drawing his patient out by reflecting the patient's statements back to him. The following conversation between a young lady and Eliza playing doctor illustrates both the Rogerian technique of encouraging a patient to keep talking and the operation of the computer program ELIZA. The first to "speak" is the young lady. The computer's responses are printed entirely in capitals.[2]

Men are all alike.

IN WHAT WAY

They're always bugging us about something or other.

CAN YOU THINK OF A SPECIFIC EXAMPLE

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 MAKE 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

DOCTOR, as ELIZA playing psychiatrist came to be known, soon became famous around the Massachusetts Institute of Technology, where it first came into existence, mainly because it was an easy program to demonstrate. Most other programs could not vividly demonstrate the information-processing power of a computer to visitors who did not already have some specialized knowledge, say,

of some branch of mathematics. DOCTOR, on the other hand, could be appreciated on some level by anyone. Its power as a demonstration vehicle was further enhanced by the fact that the visitor could actually participate in its operation. Soon copies of DOCTOR, constructed on the basis of my published description of it, began appearing at other institutions in the United States. The program became nationally known and even, in certain circles, a national plaything.

The shocks I experienced as DOCTOR became widely known and "played" were due principally to three distinct events.

1. A number of practicing psychiatrists seriously believed the DOCTOR computer program could grow into a nearly completely automatic form of psychotherapy. Colby *et al.* write, for example,

> "Further work must be done before the program will be ready for clinical use. If the 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. Because of the time-sharing capabilities of modern and future computers, several hundred patients an hour could be handled by a computer system designed for this purpose. The human therapist, involved in the design and operation of this system, would not be replaced, but would become a much more efficient man since his efforts would no longer be limited to the one-to-one patient-therapist ratio as now exists."[3]*

I had thought it essential, as a prerequisite to the very possibility that one person might help another learn to cope with his emotional problems, that the helper himself participate in the other's experience of those problems and, in large part by way of his own em-

* Nor is Dr. Colby alone in his enthusiasm for computer administered psychotherapy. Dr. Carl Sagan, the astrophysicist, recently commented on ELIZA in *Natural History,* vol. LXXXIV, no. 1 (Jan. 1975), p. 10: "No such computer program is adequate for psychiatric use today, but the same can be remarked about some human psychotherapists. In a period when more and more people in our society seem to be in need of psychiatric counseling, and when time sharing of computers is widespread, I can imagine the development of a network of computer psychotherapeutic terminals, something like arrays of large telephone booths, in which, for a few dollars a session, we would be able to talk with an attentive, tested, and largely nondirective psychotherapist."

pathic recognition of them, himself come to understand them. There are undoubtedly many techniques to facilitate the therapist's imaginative projection into the patient's inner life. But that it was possible for even one practicing psychiatrist to advocate that this crucial component of the therapeutic process be entirely supplanted by pure technique—*that* I had not imagined! What must a psychiatrist who makes such a suggestion think he is doing while treating a patient, that he can view the simplest mechanical parody of a single interviewing technique as having captured anything of the essence of a human encounter? Perhaps Colby *et al.* give us the required clue when they write;

> "A human therapist can be viewed as an information processor and decision maker with a set of decision rules which are closely linked to short-range and long-range goals, . . . He is guided in these decisions by rough empiric rules telling him what is appropriate to say and not to say in certain contexts. To incorporate these processes, to the degree possessed by a human therapist, in the program would be a considerable undertaking, but we are attempting to move in this direction."[4]

What can the psychiatrist's image of his patient be when he sees himself, as therapist, not as an engaged human being acting as a healer, but as an information processor following rules, etc.?

Such questions were my awakening to what Polanyi had earlier called a "scientific outlook that appeared to have produced a mechanical conception of man."

2. I was startled to see how quickly and how very deeply people conversing with DOCTOR became emotionally involved with the computer and how unequivocally they anthropomorphized it. Once my secretary, who had watched me work on the program for many months and therefore surely knew it to be merely a computer program, started conversing with it. After only a few interchanges with it, she asked me to leave the room. Another time, I suggested I might rig the system so that I could examine all conversations anyone had had with it, say, overnight. I was promptly bombarded with accusations that what I proposed amounted to spying on people's most

intimate thoughts; clear evidence that people were conversing with the computer as if it were a person who could be appropriately and usefully addressed in intimate terms. I knew of course that people form all sorts of emotional bonds to machines, for example, to musical instruments, motorcycles, and cars. And I knew from long experience that the strong emotional ties many programmers have to their computers are often formed after only short exposures to their machines. What I had not realized is that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people. This insight led me to attach new importance to questions of the relationship between the individual and the computer, and hence to resolve to think about them.

3. Another widespread, and to me surprising, reaction to the ELIZA program was the spread of a belief that it demonstrated a general solution to the problem of computer understanding of natural language. In my paper, I had tried to say that no general solution to that problem was possible, i.e., that language is understood only in contextual frameworks, that even these can be shared by people to only a limited extent, and that consequently even people are not embodiments of any such general solution. But these conclusions were often ignored. In any case, ELIZA was such a small and simple step. Its contribution was, if any at all, only to vividly underline what many others had long ago discovered, namely, the importance of context to language understanding. The subsequent, much more elegant, and surely more important work of Winograd[5] in computer comprehension of English is currently being misinterpreted just as ELIZA was. This reaction to ELIZA showed me more vividly than anything I had seen hitherto the enormously exaggerated attributions an even well-educated audience is capable of making, even strives to make, to a technology it does not understand. Surely, I thought, decisions made by the general public about emergent technologies depend much more on what that public attributes to such technologies than on what they actually are or can and cannot do. If, as appeared to be the case, the public's attributions are wildly misconceived, then public decisions are bound to be misguided and

often wrong. Difficult questions arise out of these observations; what, for example, are the scientist's responsibilities with respect to making his work public? And to whom (or what) is the scientist responsible?

As perceptions of these kinds began to reverberate in me, I thought, as perhaps Polanyi did after his encounter with Bukharin, that the questions and misgivings that had so forcefully presented themselves to me could be disposed of quickly, perhaps in a short, serious article. I did in fact write a paper touching on many points mentioned here.[6] But gradually I began to see that certain quite fundamental questions had infected me more chronically than I had first perceived. I shall probably never be rid of them.

There are as many ways to state these basic questions as there are starting points for coping with them. At bottom they are about nothing less than man's place in the universe. But I am professionally trained only in computer science, which is to say (in all seriousness) that I am extremely poorly educated; I can mount neither the competence, nor the courage, not even the chutzpah, to write on the grand scale actually demanded. I therefore grapple with questions that couple more directly to the concerns I have expressed, and hope that their larger implications will emerge spontaneously.

I shall thus have to concern myself with the following kinds of questions:

1. What is it about the computer that has brought the view of man as a machine to a new level of plausibility? Clearly there have been other machines that imitated man in various ways, e.g., steam shovels. But not until the invention of the digital computer have there been machines that could perform intellectual functions of even modest scope; i.e., machines that could in any sense be said to be intelligent. Now "artificial intelligence" (AI) is a subdiscipline of computer science. This new field will have to be discussed. Ultimately a line dividing human and machine intelligence must be drawn. If there is no such line, then advocates of computerized psychotherapy may be merely heralds of an age in which man has finally been recognized as nothing but a clock-work. Then the consequences of such a reality would need urgently to be divined and contemplated.

2. The fact that individuals bind themselves with strong emotional ties to machines ought not in itself to be surprising. The instruments man uses become, after all, extensions of his body. Most importantly, man must, in order to operate his instruments skillfully, internalize aspects of them in the form of kinesthetic and perceptual habits. In that sense at least, his instruments become literally part of him and modify him, and thus alter the basis of his affective relationship to himself. One would expect man to cathect more intensely to instruments that couple directly to his own intellectual, cognitive, and emotive functions than to machines that merely extend the power of his muscles. Western man's entire milieu is now pervaded by complex technological extensions of his every functional capacity. Being the enormously adaptive animal he is, man has been able to accept as authentically natural (that is, as given by nature) such technological bases for his relationship to himself, for his identity. Perhaps this helps to explain why he does not question the appropriateness of investing his most private feelings in a computer. But then, such an explanation would also suggest that the computing machine represents merely an extreme extrapolation of a much more general technological usurpation of man's capacity to act as an autonomous agent in giving meaning to his world. It is therefore important to inquire into the wider senses in which man has come to yield his own autonomy to a world viewed as machine.

3. It is perhaps paradoxical that just, when in the deepest sense man has ceased to believe in—let alone to trust—his own autonomy, he has begun to rely on autonomous machines, that is, on machines that operate for long periods of time entirely on the basis of their own internal realities. If his reliance on such machines is to be based on something other than unmitigated despair or blind faith, he must explain to himself what these machines do and even how they do what they do. This requires him to build some conception of their internal "realities." Yet most men don't understand computers to even the slightest degree. So, unless they are capable of very great skepticism (the kind we bring to bear while watching a stage magi-

cian), they can explain the computer's intellectual feats only by bringing to bear the single analogy available to them, that is, their model of their own capacity to think. No wonder, then, that they overshoot the mark; it is truly impossible to imagine a human who could imitate ELIZA, for example, but for whom ELIZA's language abilities were his limit. Again, the computing machine is merely an extreme example of a much more general phenomenon. Even the breadth of connotation intended in the ordinary usage of the word "machine," large as it is, is insufficient to suggest its true generality. For today when we speak of, for example, bureaucracy, or the university, or almost any social or political construct, the image we generate is all too often that of an autonomous machine-like process.

These, then, are the thoughts and questions which have refused to leave me since the deeper significances of the reactions to ELIZA I have described began to become clear to me. Yet I doubt that they could have impressed themselves on me as they did were it not that I was (and am still) deeply involved in a concentrate of technological society as a teacher in the temple of technology that is the Massachusetts Institute of Technology, an institution that proudly boasts of being "polarized around science and technology." There I live and work with colleagues, many of whom trust only modern science to deliver reliable knowledge of the world. I confer with them on research proposals to be made to government agencies, especially to the Department of "Defense." Sometimes I become more than a little frightened as I contemplate what we lead ourselves to propose, as well as the nature of the arguments we construct to support our proposals. Then, too, I am constantly confronted by students, some of whom have already rejected all ways but the scientific to come to know the world, and who seek only a deeper, more dogmatic indoctrination in that faith (although that word is no longer in their vocabulary). Other students suspect that not even the entire collection of machines and instruments at M.I.T. can significantly help give meaning to their lives. They sense the presence of a dilemma in an education polarized around science and technology, an education that implicitly claims to open a privileged

access-path to fact, but that cannot tell them how to decide what is to count as fact. Even while they recognize the genuine importance of learning their craft, they rebel at working on projects that appear to address themselves neither to answering interesting questions of fact nor to solving problems in theory.

Such confrontations with my own day-to-day social reality have gradually convinced me that my experience with ELIZA was symptomatic of deeper problems. The time would come, I was sure, when I would no longer be able to participate in research proposal conferences, or honestly respond to my students' need for therapy (yes, that is the correct word), without first attempting to make sense of the picture my own experience with computers had so sharply drawn for me.

Of course, the introduction of computers into our already highly technological society has, as I will try to show, merely reinforced and amplified those antecedent pressures that have driven man to an ever more highly rationalistic view of his society and an ever more mechanistic image of himself. It is therefore important that I construct my discussion of the impact of the computer on man and his society so that it can be seen as a particular kind of encoding of a much larger impact, namely, that on man's role in the face of technologies and techniques he may not be able to understand and control. Conversations around that theme have been going on for a long time. And they have intensified in the last few years.

Certain individuals of quite differing minds, temperaments, interests, and training have—however much they differ among themselves and even disagree on many vital questions—over the years expressed grave concern about the conditions created by the unfettered march of science and technology; among them are Mumford, Arendt, Ellul, Roszak, Comfort, and Boulding. The computer began to be mentioned in such discussions only recently. Now there are signs that a full-scale debate about the computer is developing. The contestants on one side are those who, briefly stated, believe computers can, should, and will do everything, and on the other side those who, like myself, believe there are limits to what computers ought to be put to do.

It may appear at first glance that this is an in-house debate of

little consequence except to a small group of computer technicians. But at bottom, no matter how it may be disguised by technological jargon, the question is whether or not every aspect of human thought is reducible to a logical formalism, or, to put it into the modern idiom, whether or not human thought is entirely computable. That question has, in one form or another, engaged thinkers in all ages. Man has always striven for principles that could organize and give sense and meaning to his existence. But before modern science fathered the technologies that reified and concretized its otherwise abstract systems, the systems of thought that defined man's place in the universe were fundamentally juridicial. They served to define man's obligations to his fellow men and to nature. The Judaic tradition, for example, rests on the idea of a contractual relationship between God and man. This relationship must and does leave room for autonomy for both God and man, for a contract is an agreement willingly entered into by parties who are free not to agree. Man's autonomy and his corresponding responsibility is a central issue of all religious systems. The spiritual cosmologies engendered by modern science, on the other hand, are infected with the germ of logical necessity. They, except in the hands of the wisest scientists and philosophers, no longer content themselves with explanations of appearances, but claim to say how things actually are and must necessarily be. In short, they convert truth to provability.

As one consequence of this drive of modern science, the question, "What aspects of life are formalizable?" has been transformed from the moral question, "How and in what form may man's obligations and responsibilities be known?" to the question, "Of what technological genus is man a species?" Even some philosophers whose every instinct rebels against the idea that man is entirely comprehensible as a machine have succumbed to this spirit of the times. Hubert Dreyfus, for example, trains the heavy guns of phenomenology on the computer model of man.[7] But he limits his argument to the technical question of what computers can and cannot do. I would argue that if computers could imitate man in every respect—which in fact they cannot—even then it would be appropriate, nay, urgent, to examine the computer in the light of man's perennial need to find his place in the world. The outcomes of practical matters that are of vital importance to everyone hinge on how and in what terms the discussion is carried out.

One position I mean to argue appears deceptively obvious: it is simply that there are important differences between men and machines as thinkers. I would argue that, however intelligent machines may be made to be, there are some acts of thought that *ought* to be attempted only by humans. One socially significant question I thus intend to raise is over the proper place of computers in the social order. But, as we shall see, the issue transcends computers in that it must ultimately deal with logicality itself—quite apart from whether logicality is encoded in computer programs or not.

The lay reader may be forgiven for being more than slightly incredulous that anyone should maintain that human thought is entirely computable. But his very incredulity may itself be a sign of how marvelously subtly and seductively modern science has come to influence man's imaginative construction of reality.

Surely, much of what we today regard as good and useful, as well as much of what we would call knowledge and wisdom, we owe to science. But science may also be seen as an addictive drug. Not only has our unbounded feeding on science caused us to become dependent on it, but, as happens with many other drugs taken in increasing dosages, science has been gradually converted into a slow-acting poison. Beginning perhaps with Francis Bacon's misreading of the genuine promise of science, man has been seduced into wishing and working for the establishment of an age of rationality, but with his vision of rationality tragically twisted so as to equate it with logicality. Thus have we very nearly come to the point where almost every genuine human dilemma is seen as a mere paradox, as a merely apparent contradiction that could be untangled by judicious applications of cold logic derived from a higher standpoint. Even murderous wars have come to be perceived as mere problems to be solved by hordes of professional problemsolvers. As Hannah Arendt said about recent makers and executors of policy in the Pentagon:

> "They were not just intelligent, but prided themselves on being 'rational' . . . They were eager to find formulas, preferably expressed in a pseudo-mathematical language, that would unify the

> most disparate phenomena with which reality presented them; that is, they were eager to discover *laws* by which to explain and predict political and historical facts as though they were as necessary, and thus as reliable, as the physicists once believed natural phenomena to be . . . [They] did not *judge;* they calculated. . . . an utterly irrational confidence in the calculability of reality [became] the leitmotif of the decision making."[8]

And so too have nearly all political confrontations, such as those between races and those between the governed and their governors, come to be perceived as mere failures of communication. Such rips in the social fabric can then be systematically repaired by the expert application of the latest information-handling techniques—at least so it is believed. And so the rationality-is-logicality equation, which the very success of science has drugged us into adopting as virtually an axiom, has led us to deny the very existence of human conflict, hence the very possibility of the collision of genuinely incommensurable human interests and of disparate human values, hence the existence of human values themselves.

It may be that human values are illusory, as indeed B. F. Skinner argues. If they are, then it is presumably up to science to demonstrate that fact, as indeed Skinner (as scientist) attempts to do. But then science must itself be an illusory system. For the only certain knowledge science can give us is knowledge of the behavior of formal systems, that is, systems that are games invented by man himself and in which to assert truth is nothing more or less than to assert that, as in a chess game, a particular board position was arrived at by a sequence of legal moves. When science purports to make statements about man's experiences, it bases them on identifications between the primitive (that is, undefined) objects of one of its formalisms, the pieces of one of its games, and some set of human observations. No such sets of correspondences can ever be proved to be correct. At best, they can be falsified, in the sense that formal manipulations of a system's symbols may lead to symbolic configurations which, when read in the light of the set of correspondences in question, yield interpretations contrary to empirically observed phenomena. Hence all empirical science is an elaborate structure built on piles that are anchored, not on bedrock as is commonly

supposed, but on the shifting sand of fallible human judgment, conjecture, and intuition. It is not even true, again contrary to common belief, that a single purported counter-instance that, if accepted as genuine would certainly falsify a specific scientific theory, generally leads to the immediate abandonment of that theory. Probably all scientific theories currently accepted by scientists themselves (excepting only those purely formal theories claiming no relation to the empirical world) are today confronted with contradicting evidence of more than negligible weight that, again if fully credited, would logically invalidate them. Such evidence is often explained (that is, explained away) by ascribing it to error of some kind, say, observational error, or by characterizing it as inessential, or by the assumption (that is, the faith) that some yet-to-be-discovered way of dealing with it will some day permit it to be acknowledged but nevertheless incorporated into the scientific theories it was originally thought to contradict. In this way scientists continue to rely on already impaired theories and to infer "scientific fact" from them.*

The man in the street surely believes such scientific facts to be as well-established, as well-proven, as his own existence. His certitude is an illusion. Nor is the scientist himself immune to the same illusion. In his praxis, he must, after all, suspend disbelief in order to do or think anything at all. He is rather like a theatergoer, who, in order to participate in and understand what is happening on the stage, must for a time pretend to himself that he is witnessing real events. The scientist must believe his working hypothesis, together with its vast underlying structure of theories and assumptions, even if only for the sake of the argument. Often the "argument" extends over his entire lifetime. Gradually he becomes what he at first merely pretended to be: a true believer. I choose the word "argument" thoughtfully, for scientific demonstrations, even mathematical proofs, are fundamentally acts of persuasion.

---

* Thus, Charles Everett writes on the now-discarded phlogiston theory of combustion (in the *Encyclopaedia Britannica*, 11th ed., 1911, vol. VI, p. 34): "The objections of the anti-phlogistonists, such as the fact that the calices weigh more than the original metals instead of less as the theory suggests, were answered by postulating that phlogiston was a principle of levity, or even completely ignored as an accident, the change in qualities being regarded as the only matter of importance." Everett lists H. Cavendish and J. Priestley, both great scientists of their time, as adherents to the phlogiston theory.

Scientific statements can never be certain; they can be only more or less credible. And credibility is a term in individual psychology, i.e., a term that has meaning only with respect to an individual observer. To say that some proposition is credible is, after all, to say that it is believed by an agent who is free not to believe it, that is, by an observer who, after exercising judgment and (possibly) intuition, chooses to accept the proposition as worthy of his believing it. How then can science, which itself surely and ultimately rests on vast arrays of human value judgments, demonstrate that human value judgments are illusory? It cannot do so without forfeiting its own status as the single legitimate path to understanding man and his world.

But no merely logical argument, no matter how cogent or eloquent, can undo this reality: that science has become the sole legitimate form of understanding in the common wisdom. When I say that science has been gradually converted into a slow-acting poison, I mean that the attribution of certainty to scientific knowledge by the common wisdom, an attribution now made so nearly universally that it has become a commonsense dogma, has virtually delegitimatized all other ways of understanding. People viewed the arts, especially literature, as sources of intellectual nourishment and understanding, but today the arts are perceived largely as entertainments. The ancient Greek and Oriental theaters, the Shakespearian stage, the stages peopled by the Ibsens and Chekhovs nearer to our day—these were schools. The curricula they taught were vehicles for understanding the societies they represented. Today, although an occasional Arthur Miller or Edward Albee survives and is permitted to teach on the New York or London stage, the people hunger only for what is represented to them to be scientifically validated knowledge. They seek to satiate themselves at such scientific cafeterias as *Psychology Today,* or on popularized versions of the works of Masters and Johnson, or on scientology as revealed by L. Ron Hubbard. Belief in the rationality-logicality equation has corroded the prophetic power of language itself. We can count, but we are rapidly forgetting how to say what is worth counting and why.

10

AGAINST THE IMPERIALISM OF INSTRUMENTAL REASON

That man has aggregated to himself enormous power by means of his science and technology is so grossly banal a platitude that, paradoxically, although it is as widely believed as ever, it is less and less often repeated in serious conversation. The paradox arises because a platitude that ceases to be commonplace ceases to be perceived as a platitude. Some circles may even, after it has not been heard for a while, perceive it as its very opposite, that is, as a deep truth. There is a parable in that, too: the power man has acquired through his science and technology has itself been converted into impotence.

The common people surely feel this. Studs Terkel, in a monumental study of daily work in America, writes;

> "For the many there is hardly concealed discontent. . . . 'I'm a machine,' says the spot welder. 'I'm caged,' says the bank teller,

> and echoes the hotel clerk. 'I'm a mule,' says the steel worker. 'A monkey can do what I do,' says the receptionist. 'I'm less than a farm implement,' says the migrant worker. 'I'm an object,' says the high fashion model. Blue collar and white call upon the identical phrase: 'I'm a robot.'"[1]

Perhaps the common people believe that, although they are powerless, there is power, namely, that exercised by their leaders. But we have seen that the American Secretary of State believes that events simply "befall" us, and that the American Chief of the Joint Chiefs of Staff confesses to having become a slave of computers. Our leaders cannot find the power either.

Even physicians, formerly a culture's very symbol of power, are powerless as they increasingly become mere conduits between their patients and the major drug manufacturers. Patients, in turn, are more and more merely passive objects on whom cures are wrought and to whom things are done. Their own inner healing resources, their capacities for self-reintegration, whether psychic or physical, are more and more regarded as irrelevant in a medicine that can hardly distinguish a human patient from a manufactured object. The now ascendant biofeedback movement may be the penultimate act in the drama separating man from nature; man no longer even senses himself, his body, directly, but only through pointer readings, flashing lights, and buzzing sounds produced by instruments attached to him as speedometers are attached to automobiles. The ultimate act of the drama is, of course, the final holocaust that wipes life out altogether.

Technological inevitability can thus be seen to be a mere element of a much larger syndrome. Science promised man power. But, as so often happens when people are seduced by promises of power, the price exacted in advance and all along the path, and the price actually paid, is servitude and impotence. Power is nothing if it is not the power to choose. Instrumental reason can make decisions, but there is all the difference between deciding and choosing.

The people Studs Terkel is talking about make decisions all day long, every day. But they appear not to make choices. They are, as they themselves testify, like Winograd's robot. One asks it "Why did you do that?" and it answers "Because this or that decision

branch in my program happened to come out that way." And one asks "Why did you get to that branch?" and it again answers in the same way. But its final answer is "Because you told me to." Perhaps every human act involves a chain of calculations at what a systems engineer would call decision nodes. But the difference between a mechanical act and an authentically human one is that the latter terminates at a node whose decisive parameter is not "Because you told me to," but "Because I chose to." At that point calculations and explanations are displaced by truth. Here, too, is revealed the poverty of Simon's hypothesis that

> "The whole man, like the ant, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself."

For that hypothesis to be true, it would also have to be true that man's capacity for choosing is as limited as is the ant's, that man has no more will or purpose, and, perhaps most importantly, no more a self-transcendent sense of obligation to himself as part of the continuum of nature, than does the ant. Again, it is a mystery why anyone would want to believe this to be the true condition of man.

But now and then a small light appears to penetrate the murky fog that obscures man's authentic capacities. Recently, for example, a group of eminent biologists urged their colleagues to discontinue certain experiments in which new types of biologically functional bacterial plasmids are created.[2] They express "serious concern that some of these artificial recombinant DNA molecules could prove biologically hazardous." Their concern is, so they write, "for the possible unfortunate consequences of the indiscriminate application of these techniques." Theirs is certainly a step in the right direction, and their initiative is to be applauded. Still, one may ask, why do they feel they have to give a reason for what they recommend at all? Is not the overriding obligation on men, including men of science, to exempt life itself from the madness of treating everything as an object, a sufficient reason, and one that does not even have to be spoken? Why does it have to be explained? It would

appear that even the noblest acts of the most well-meaning people are poisoned by the corrosive climate of values of our time.

An easy explanation of this, and perhaps it contains truth, is that well-meaningness has supplanted nobility altogether. But there is a more subtle one. Our time prides itself on having finally achieved the freedom from censorship for which libertarians in all ages have struggled. Sexual matters can now be discussed more freely than ever before, women are beginning to find their rightful place in society, and, in general, ideas that could be only whispered until a decade or so ago may now circulate without restriction. The credit for these great achievements is claimed by the new spirit of rationalism, a rationalism that, it is argued, has finally been able to tear from man's eyes the shrouds imposed by mystical thought, religion, and such powerful illusions as freedom and dignity. Science has given to us this great victory over ignorance. But, on closer examination, this victory too can be seen as an Orwellian triumph of an even higher ignorance: what we have gained is a new conformism, which permits us to say anything that can be said in the functional languages of instrumental reason, but forbids us to allude to what Ionesco called the living truth. Just as our television screens may show us unbridled violence in "living color" but not scenes of authentic intimate love—the former by an itself-obscene reversal of values is said to be "real," whereas the latter is called obscene—so we may discuss the very manufacture of life and its "objective" manipulation, but we may not mention God, grace, or morality. Perhaps the biologists who urge their colleagues to do the right thing, but for the wrong reasons, are in fact motivated by their own deep reverence for life and by their own authentic humanity, only they dare not say so. In any case, such arguments would not be "effective," that is to say, instrumental.

If that is so, then those who censor their own speech do so, to use an outmoded expression, at the peril of their souls.

There is still another way to justify a scientist's renunciation of a particular line of research—and it is one from which all of us may derive lessons pertinent to our own lives. It begins from the principle that the range of one's responsibilities must be commensurate with the range of the effects of one's actions. In earlier times this

principle led to a system of ethics that concerned itself chiefly with how persons conducted themselves toward one another. The biblical commandments, for example, speak mainly of what an individual's duties are toward his family and his neighbors. In biblical times few people could do anything that was likely to affect others beyond the boundaries of their own living spaces. Man's science and technology have altered this circumstance drastically. Not only can modern man's actions affect the whole planet that is his habitat, but they can determine the future of the entire human species. It follows therefore that man, particularly man the scientist and engineer, has responsibilities that transcend his immediate situation, that in fact extend directly to future generations. These responsibilities are especially grave since future generations cannot advocate their own cause now. We are all their trustees.[3]

The biologists' overt renunciation, however they themselves justify it, is an example which it behooves all scientists to emulate. Is this to suggest that scientists should close their minds to certain kinds of "immoral" hypotheses? Not at all. A scientific hypothesis is, at least from a scientific point of view, either true or false. This applies, for example, to Simon's hypotheses that man is "quite simple" and that he can be entirely simulated by a machine, as well as to McCarthy's hypothesis that there exists a logical calculus in terms of which all of reality can be formalized. It would be a silly error of logic to label such (or any other) hypotheses either moral or immoral or, for that matter, responsible or irresponsible.

But, although a scientific hypothesis can itself have no moral or ethical dimensions, an individual's decision to adopt it even tentatively, let alone to announce his faith in it to the general public, most certainly involves value judgments and does therefore have such dimensions. As the Harvard economist Marc J. Roberts recently wrote,

> "Suppose we must choose between two hypotheses. No matter which we select, there is always the possibility that the other is correct. Obviously the relative likelihood of making a mistake when we select one or the other matters—but so too do the costs of alternative mistakes, the costs of assuming *A* is true when in fact *B* is true or vice versa. We might well choose to risk a more likely small cost than a less likely large one. Yet the magnitude of the cost of being wrong in each case cannot be determined except on the basis of our values.
>
> "Consider an extreme example: the view that there are genetic differences in the mental functioning of different races. Suppose society were to accept this view, and it proved false. I believe that very great evil would have been done. On the other hand, suppose society adopted the view that there are no differences, and that turned out to be incorrect. I would expect much less harm to result. Given these costs, I would want evidence which made the hypothesis of interracial similarity very unlikely indeed before I would reject it. My scientific choice depends on my values, not because I am uncritical or would like to believe that there are no such differences, but because consistent choices under uncertainty can only be made by looking at the cost of making alternative kinds of errors. In contrast, a would-be 'value-neutral scientist' would presumably be willing to operate on the assumption that such differences exist as soon as evidence made it even slightly more likely than the reverse assumption.
>
> "These questions do not arise routinely in scientific work because traditional statistical methods typically subsume them under the choice of test criteria or of the particular technique to be used in estimating some magnitude. That choice is then made on conventional or traditional grounds, usually without discussion, justification, or even acknowledgement that value choices have been made."[4]

Roberts chose to illustrate that scientific hypotheses are not "value free" by citing the values enter into the scientist's choice to tolerate or not to tolerate the potential cost of being wrong. Values, as I will try to show, enter into choices made by scientists in other (and I believe even more important) ways as well. For the moment, however, I mean only to assert that it is entirely proper to say "bravo" to the biologists whose example we have cited, and to say "shame" to the scientists who recently wrote that "a machine-animal symbiont with an animal visual system and brain to augment mechanical functions" will be technically "feasible" within the next fifteen years.[5]

The introduction of words like "ethics" and "ought" into conversations about science seems almost always to engender a tension not unlike, I would say, the strain one can sense rising whenever, in conversation with elderly German university professors, one happens to allude to the career of one of their colleagues who prospered during the Hitler years. In the latter situation, the lowering of the social temperature betrays the fear that something "unfortunate" might be said, especially that the colleague's past inability to renounce his personal ambitions for the sake of morality might be mentioned. There is a recognition, then, of course, that the conduct not only of the colleague, but of all German academicians of the time, is in question. In the former situation, the tension betrays a similar concern, for ethics, at bottom, deals with nothing so much as renunciation. The tension betrays the fear that something will be said about what science, that is, scientists, ought and ought not to do. And there is a recognition that what might be talked about doesn't apply merely to science generally or to some abstract population known as scientists, but to the very people present.

Some scientists, though by no means all, maintain that the domain of science is universal, that there can be nothing which, as a consequence of some "higher" principle, ought not to be studied. And from this premise the conclusion is usually drawn that any talk of ethical "oughts" which apply to science is inherently subversive and anti-scientific, even anti-intellectual.

Whatever the merits of this argument as abstract logic may be, it is muddleheaded when applied to concrete situations, for there are infinitely many questions open to scientific investigation, but only finite resources at the command of science. Man must therefore choose which questions to attack and which to leave aside. We don't know, for example, whether the number of pores on an individual's skin is in any way correlated with the number of neurons in his brain. There is no interest in that question, and therefore no controversy about whether or not science ought to study it. The Chinese have practiced acupuncture for many centuries without arousing the interest of Western science. Now, suddenly, Western scientists have become interested. These examples illustrate that scientific "progress" does not move along some path determined by nature itself, but that it mirrors human interests and concerns.

Surely finely honed human intelligence is among the scarcest of resources available to modern society. And clearly some problems amenable to scientific investigation are more important than others. Human society is therefore inevitably faced with the task of wisely distributing the scarce resource that is its scientific talent. There simply is a responsibility—it cannot be wished away—to decide which problems are more important or interesting or whatever than others. Every specific society must constantly find ways to meet that responsibility. The question here is *how,* in an open society, these ways are to be found; are they to be dictated by, say, the military establishment, or are they to be open to debate among citizens and scientists? If they are to be debated, then why are ethics to be excluded from the discussion? And, finally, how can anything sensible emerge unless all first agree that, contrary to what John von Neuman asserted, technological possibilities are not irresistible to man? "Can" does not imply "ought."

Unfortunately, the new conformism that permits us to speak of everything except the few simple truths that are written in our hearts and in the holy books of each of man's many religions renders all arguments based on these truths—no matter how well thought out or eloquently constructed—laughable in the eyes of the scientists and technicians to whom they may be addressed. This in itself is probably the most tragic example of how an idea, badly used, turns into its own opposite. Scientists who continue to prattle on about "knowledge for its own sake" in order to exploit that slogan for their self-serving ends have detached science and knowledge from any contact with the real world. A central question of knowledge, once won, is its validation; but what we now see in almost all fields, especially in the branches of computer science we have been discussing, is that the validation of scientific knowledge has been reduced to the display of technological wonders. This can be interpreted in one of only two ways: either the nature to which science is attached consists entirely of raw material to be molded and manipulated as an object; or the knowledge that science has purchased for

does not distinguish between basic & applied research

man is entirely irrelevant to man himself. Science cannot agree that the latter is true, for if it were, science would lose its license to practice. That loss would, of course, entail practical consequence (involving money and all that) which scientists would resist with all their might. If the former is true, then man himself has become an object. There is abundant evidence that this is, in fact, what has happened. But then knowledge too has lost the purity of which scientists boast so much; it has then become an enterprise no more or less important and no more inherently significant than, say, the knowledge of how to lay out an automobile assembly line. Who would want to know that "for its own sake"?

This development is tragic, in that it robs science of even the possibility of being guided by any authentically human standards, while it in no way restricts science's potential to deliver ever-increasing power to men. And here too we find the root of the much-talked-about dehumanization of man. An individual is dehumanized whenever he is treated as less than a whole person. The various forms of human and social engineering we have discussed here do just that, in that they circumvent all human contexts, especially those that give real meaning to human language.

The fact that arguments which appeal to higher principles—say, to an individual's obligations to his children, or to nature itself—are not acknowledged as legitimate poses a serious dilemma for anyone who wishes to persuade his colleagues to cooperate in imposing some limits on their research. If he makes such arguments anyway, perhaps hoping to induce a kind of conversion experience in his colleagues, then he risks being totally ineffective and even being excommunicated as a sort of comic fool. If he argues for restraint on the grounds that irreversible consequences may follow unrestrained research, then he participates in and helps to legitimate the abuse of instrumental reason (say, in the guise of cost-benefit analyses) against which he intends to struggle.

As is true of so many other dilemmas, the solution to this one lies in rejecting the rules of the game that give rise to it. For the present dilemma, the operative rule is that the salvation of the world—and that *is* what I am talking about—depends on converting

others to sound ideas. That rule is false. The salvation of the world depends only on the individual whose world it is. At least, every individual must act as if the whole future of the world, of humanity itself, depends on him. Anything less is a shirking of responsibility and is itself a dehumanizing force, for anything less encourages the individual to look upon himself as a mere actor in a drama written by anonymous agents, as less than a whole person, and that is the beginning of passivity and aimlessness.

This is not an argument for solipsism, nor is it a counsel for every man to live only for himself. But it does argue that every man must live for himself first. For only by experiencing his own intrinsic worth, a worth utterly independent of his "use" as an instrument, can he come to know those self-transcendent ends that ultimately confer on him his identity and that are the only ultimate validators of human knowledge.

But the fact that each individual is responsible for the whole world, and that the discharge of that responsibility involves first of all each individual's responsibility to himself, does not deny that all of us have duties to one another. Chief among these is that we instruct one another as best we can. And the principal and most effective form of instruction we can practice is the example our own conduct provides to those who are touched by it. Teachers and writers have an especially heavy responsibility, precisely because they have taken positions from which their example reaches more than the few people in their immediate circle.

This spirit dictates that I must exhibit some of my own decisions about what I may and may not do in computer science. I do so with some misgivings, for I have learned that people are constantly asking one another what they must do, whereas the only really important question is what they must be. The physicist Steven Weinberg, in commenting on recent criticisms of science, writes, for example,

> "I have tried to understand these critics by looking through some of their writings, and have found a good deal that is perti-

> nent, and even moving. I especially share their distrust of those, from David Ricardo to the Club of Rome, who too confidently apply the methods of the natural sciences to human affairs. But in the end I am puzzled. What is it they want *me* to do?"[6]

My fear is that I will be understood to be answering a question of the kind Weinberg asks. That is not my intention. But the risk that I will be misunderstood cannot excuse me from my duty.

There is, in my view, no project in computer science as such that is morally repugnant and that I would advise students or colleagues to avoid. The projects I have been discussing, and others I will mention, are not properly part of computer science. Computers are not central to the work of Forrester and Skinner. The others are not computer science, because they are for the most part not science at all. They are, as I have already suggested, clever aggregations of techniques aimed at getting something done. Perhaps because of the accidents of history that caused academic departments whose concerns are with computers to be called "computer science" departments, all work done in such departments is indiscriminately called "science," even if only part of it deserves that honorable appellation. Tinkerers with techniques (gadget worshippers, Norbert Wiener called them) sometimes find it hard to resist the temptation to associate themselves with science and to siphon legitimacy from the reservoir it has accumulated. But not everyone who calls himself a singer has a voice.

Not all projects, by very far, that are frankly performance-oriented are dangerous or morally repugnant. Many really do help man to carry on his daily work more safely and more effectively. Computer-controlled navigation and collision-avoidance devices, for example, enable ships and planes to function under hitherto disabling conditions. The list of ways in which the computer has proved helpful is undoubtedly long. There are, however, two kinds of computer applications that either ought not be undertaken at all, or, if they are contemplated, should be approached with utmost caution.

The first kind I would call simply obscene. These are ones whose very contemplation ought to give rise to feelings of disgust in

every civilized person. The proposal I have mentioned, that an animal's visual system and brain be coupled to computers, is an example. It represents an attack on life itself. One must wonder what must have happened to the proposers' perception of life, hence to their perceptions of themselves as part of the continuum of life, that they can even think of such a thing, let alone advocate it. On a much lesser level, one must wonder what conceivable need of man could be fulfilled by such a "device" at all, let alone by only such a device.

I would put all projects that propose to substitute a computer system for a human function that involves interpersonal respect, understanding, and love in the same category. I therefore reject Colby's proposal that computers be installed as psychotherapists, not on the grounds that such a project might be technically infeasible, but on the grounds that it is immoral. I have heard the defense that a person may get some psychological help from conversing with a computer even if the computer admittedly does not "understand" the person. One example given me was of a computer system designed to accept natural-language text via its typewriter console, and to respond to it with a randomized series of "yes" and "no." A troubled patient "conversed" with this system, and was allegedly led by it to think more deeply about his problems and to arrive at certain allegedly helpful conclusions. Until then he had just drifted in aimless worry. In principle, a set of Chinese fortune cookies or a deck of cards could have done the same job. The computer, however, contributed a certain aura—derived, of course, from science—that permitted the "patient" to believe in it where he might have dismissed fortune cookies and playing cards as instruments of superstition. The question then arises, and it answers itself, do we wish to encourage people to lead their lives on the basis of patent fraud, charlatanism, and unreality? And, more importantly, do we really believe that it helps people living in our already overly machine-like world to prefer the therapy administered by machines to that given by other people? I have heard this latter question answered with the assertion that my position is nothing more than "let them eat cake." It is said to ignore the shortage of good human psychotherapists, and to deny to troubled people what little help computers can now give them merely because presently available computers don't "yet"

measure up to, say, the best psychoanalysis. But that objection misses the point entirely. The point is (Simon and Colby to the contrary notwithstanding) that there are some human functions for which computers *ought* not to be substituted. It has nothing to do with what computers can or cannot be made to do. Respect, understanding, and love are not technical problems.

The second kind of computer application that ought to be avoided, or at least not undertaken without very careful forethought, is that which can easily be seen to have irreversible and not entirely foreseeable side effects. If, in addition, such an application cannot be shown to meet a pressing human need that cannot readily be met in any other way, then it ought not to be pursued. The latter stricture follows directly from the argument I have already presented about the scarcity of human intelligence.

The example I wish to cite here is that of the automatic recognition of human speech. There are now three or four major projects in the United States devoted to enabling computers to understand human speech, that is, to programming them in such a way that verbal speech directed at them can be converted into the same internal representations that would result if what had been said to them had been typed into their consoles.

The problem, as can readily be seen, is very much more complicated than that of natural-language understanding as such, for in order to understand a stream of coherent speech, the language in which that speech is rendered must be understood in the first place. The solution of the "speech-understanding problem" therefore presupposes the solution of the "natural-language-understanding problem." And we have seen that, for the latter, we have only "the tiniest bit of relevant knowledge." But I am not here concerned with the technical feasibility of the task, nor with any estimate of just how little or greatly optimistic we might be about its completion.

Why should we want to undertake this task at all? I have asked this question of many enthusiasts for the project. The most cheerful answer I have been able to get is that it will help physicians record their medical notes and then translate these notes into action more efficiently. Of course, anything that has any ostensible connection to medicine is automatically considered good. But here we have to remember that the problem is so enormous that only the largest possible computers will ever be able to manage it. In other words, even if the desired system were successfully designed, it would probably require a computer so large and therefore so expensive that only the largest and best-endowed hospitals could possibly afford it—but in fact the whole system might be so prohibitively expensive that even they could not afford it. The question then becomes, is this really what medicine needs most at this time? Would not the talent, not to mention the money and the resources it represents, be better spent on projects that attack more urgent and more fundamental problems of health care?

But then, this alleged justification of speech-recognition "research" is merely a rationalization anyway. (I put the word "research" in quotation marks because the work I am here discussing is mere tinkering. I have no objection to serious scientists studying the psycho-physiology of human speech recognition.) If one asks such questions of the principal sponsor of this work, the Advanced Research Projects Agency (ARPA) of the United States Department of Defense, as was recently done at an open meeting, the answer given is that the Navy hopes to control its ships, and the other services their weapons, by voice commands. This project then represents, in the eyes of its chief sponsor, a long step toward a fully automated battlefield. I see no reason to advise my students to lend their talents to that aim.

I have urged my students and colleagues to ask still another question about this project: Granted that a speech-recognition machine is bound to be enormously expensive, and that only governments and possibly a very few very large corporations will therefore be able to afford it, what will they use it for? What can it possibly be used for? There is no question in my mind that there is no pressing human problem that will more easily be solved because such machines exist. But such listening machines, could they be made, will make monitoring of voice communication very much easier than it now is. Perhaps the only reason that there is very little government surveillance of telephone conversations in many countries of the world is that such surveillance takes so much manpower. Each con-

versation on a tapped phone must eventually be listened to by a human agent. But speech-recognizing machines could delete all "uninteresting" conversations and present transcripts of only the remaining ones to their masters. I do not for a moment believe that we will achieve this capability within the future so clearly visible to Newell and Simon. But I do ask, why should a talented computer technologist lend his support to such a project? As a citizen I ask, why should my government spend approximately 2.5 million dollars a year (as it now does) on this project?

Surely such questions presented themselves to thoughtful people in earlier stages of science and technology. But until recently society could always meet the unwanted and dangerous effects of its new inventions by, in a sense, reorganizing itself to undo or to minimize these effects. The density of cities could be reduced by geographically expanding the city. An individual could avoid the terrible effects of the industrial revolution in England by moving to America. And America could escape many of the consequences of the increasing power of military weapons by retreating behind its two oceanic moats. But those days are gone. The scientist and the technologist can no longer avoid the responsibility for what he does by appealing to the infinite powers of society to transform itself in response to new realities and to heal the wounds he inflicts on it. Certain limits have been reached. The transformations the new technologies may call for may be impossible to achieve, and the failure to achieve them may mean the annihilation of all life. No one has the right to impose such a choice on mankind.

I have spoken here of what ought and ought not to be done, of what is morally repugnant, and of what is dangerous. I am, of course, aware of the fact that these judgments of mine have themselves no moral force except on myself. Nor, as I have already said, do I have any intention of telling other people what tasks they should and should not undertake. I urge them only to consider the consequences of what they do do. And here I mean not only, not even primarily, the direct consequences of their actions on the world about them. I mean rather the consequences on themselves, as they construct their rationalizations, as they repress the truths that urge them to different courses, and as they chip away at their own autonomy. That so many people so often ask what they must do is a sign that the order of being and doing has become inverted. Those who know who and what they are do not need to ask what they should do. And those who must ask will not be able to stop asking until they begin to look inside themselves. But it is everyone's task to show by example what questions one can ask of oneself, and to show that one can live with what few answers there are.

But just as I have no license to dictate the actions of others, neither do the constructors of the world in which I must live have a right to unconditionally impose their visions on me. Scientists and technologists have, because of their power, an especially heavy responsibility, one that is not to be sloughed off behind a facade of slogans such as that of technological inevitability. In a world in which man increasingly meets only himself, and then only in the form of the products he has made, the makers and designers of these products—the buildings, airplanes, foodstuffs, bombs, and so on—need to have the most profound awareness that their products are, after all, the results of human choices. Men could instead choose to have truly safe automobiles, decent television, decent housing for everyone, or comfortable, safe, and widely distributed mass transportation. The fact that these things do not exist, in a country that has the resources to produce them, is a consequence, not of technological inevitability, not of the fact that there is no longer anyone who makes choices, but of the fact that people have chosen to make and to have just exactly the things we have made and do have.

It is hard, when one sees a particularly offensive television commercial, to imagine that adult human beings sometime and somewhere sat around a table and decided to construct exactly that commercial and to have it broadcast hundreds of times. But that is what happens. These things are not products of anonymous forces. They are the products of groups of men who have agreed among themselves that this pollution of the consciousness of the people serves their purposes.

But, as has been true since the beginning of recorded history, decisions having the most evil consequences are often made in the service of some overriding good. For example, in the summer of 1966 there was considerable agitation in the United States over America's

intensive bombing of North Viet Nam. (The destruction rained on South Viet Nam by American bombers was less of an issue in the public debate, because the public was still persuaded that America was "helping" that unfortunate land.) Approximately forty American scientists who were high in the scientific estate decided to help stop the bombing by convening a summer study group under the auspices of the Institute of Defense Analyses, a prestigious consulting firm for the Department of Defense. They intended to demonstrate that the bombing was in fact ineffective.[7]

They made their demonstration using the best scientific tools, operations research and systems analysis and all that. But they felt they would not be heard by the Secretary of Defense unless they suggested an alternative to the bombing. They proposed that an "electronic fence" be placed in the so-called demilitarized zone separating South from North Viet Nam. This barrier was supposed to stop infiltrators from the North. It was to consist of, among other devices, small mines seeded into the earth, and specifically designed to blow off porters' feet but to be insensitive to truck passing over them. Other devices were to interdict truck traffic. The various electronic sensors, their monitors, and so on, eventually became part of the so-called McNamara line. This was the beginning of what has since developed into the concept of the electronic battlefield.

The intention of most of these men was not to invent or recommend a new technology that would make warfare more terrible and, by the way, less costly to highly industrialized nations at the expense of "underdeveloped" ones. Their intention was to stop the bombing. In this they were wholly on the side of the peace groups and of well-meaning citizens generally. And they actually accomplished their objective; the bombing of North Viet Nam was stopped for a time and the McNamara fence was installed. However, these enormously visible and influential people could have instead simply announced that they believed the bombing, indeed the whole American Viet Nam adventure, to be wrong, and that they would no longer "help." I know that at least some of the participants believed that the war was wrong; perhaps all of them did. But, as some of them explained to me later, they felt that if they made such an announcement, they would not be listened to, then or ever again.

Yet, who can tell what effect it would have had if forty of America's leading scientists had, in the summer of 1966, joined the peace groups in coming out flatly against the war on moral grounds? Apart from the positive effect such a move might have had on world events, what negative effect did their compromise have on themselves and on their colleagues and students for whom they served as examples?

There are several lessons to be learned from this episode. The first is that it was not technological inevitability that invented the electronic battlefield, nor was it a set of anonymous forces. Men just like the ones who design television commercials sat around a table and chose. Yet the outcome of the debates of the 1966 Summer Study were in a sense foreordained. The range of answers one gets is determined by the domain of questions one asks. As soon as it was settled that the Summer Study was to concern itself with only technical questions, the solution to the problem of stopping the bombing of the North became essentially a matter of calculation. When the side condition was added that the group must at all costs maintain its credibility with its sponsors, that it must not imperil the participants' "insider" status, then all degrees of freedom that its members might have had initially were effectively lost. Many of the participants have, I know, defended academic freedom, their own as well as that of colleagues whose careers were in jeopardy for political reasons. These men did not perceive themselves to be risking their scholarly or academic freedoms when they engaged in the kind of consulting characterized by the Summer Study. But the sacrifice of the degrees of freedom they might have had if they had not so thoroughly abandoned themselves to their sponsors, whether they made that sacrifice unwittingly or not, was a more potent form of censorship than any that could possibly have been imposed by officials of the state. This kind of intellectual self-mutilation, precisely because it is largely unconscious, is a principal source of the feeling of powerlessness experienced by so many people who appear, superficially at least, to occupy seats of power.

A second lesson is this. These men were able to give the counsel they gave because they were operating at an enormous psychological distance from the people who would be maimed and

killed by the weapons systems that would result from the ideas they communicated to their sponsors. The lesson, therefore, is that the scientist and technologist must, by acts of will and of the imagination, actively strive to reduce such psychological distances, to counter the forces that tend to remove him from the consequences of his actions. He must—it is as simple as this—think of what he is actually doing. He must learn to listen to his own inner voice. He must learn to say "No!"

Finally, it is the act itself that matters. When instrumental reason is the sole guide to action, the acts it justifies are robbed of their inherent meanings and thus exist in an ethical vacuum. I recently heard an officer of a great university publicly defend an important policy decision he had made, one that many of the university's students and faculty opposed on moral grounds, with the words: "We could have taken a moral stand, but what good would that have done?" But the good of a moral act inheres in the act itself. That is why an act can itself ennoble or corrupt the person who performs it. The victory of instrumental reason in our time has brought about the virtual disappearance of this insight and thus perforce the delegitimation of the very idea of nobility.

I am aware, of course, that hardly anyone who reads these lines will feel himself addressed by them—so deep has the conviction that we are all governed by anonymous forces beyond our control penetrated into the shared consciousness of our time. And accompanying this conviction is a debasement of the idea of civil courage.

It is a widely held but a grievously mistaken belief that civil courage finds exercise only in the context of world-shaking events. To the contrary, its most arduous exercise is often in those small contexts in which the challenge is to overcome the fears induced by petty concerns over career, over our relationships to those who appear to have power over us, over whatever may disturb the tranquility of our mundane existence.

If this book is to be seen as advocating anything, then let it be a call to this simple kind of courage. And, because this book is, after all, about computers, let that call be heard mainly by teachers of computer science.

I want them to have heard me affirm that the computer is a powerful new metaphor for helping us to understand many aspects of the world, but that it enslaves the mind that has no other metaphors and few other resources to call on. The world is many things, and no single framework is large enough to contain them all, neither that of man's science nor that of his poetry, neither that of calculating reason nor that of pure intuition. And just as a love of music does not suffice to enable one to play the violin—one must also master the craft of the instrument and of music itself—so is it not enough to love humanity in order to help it survive. The teacher's calling to teach his craft is therefore an honorable one. But he must do more than that: he must teach more than one metaphor, and he must teach more by the example of his conduct than by what he writes on the blackboard. He must teach the limitations of his tools as well as their power.

It happens that programming is a relatively easy craft to learn. Almost anyone with a reasonably orderly mind can become a fairly good programmer with just a little instruction and practice. And because programming is almost immediately rewarding, that is, because a computer very quickly begins to behave somewhat in the way the programmer intends it to, programming is very seductive, especially for beginners. Moreover, it appeals most to precisely those who do not yet have sufficient maturity to tolerate long delays between an effort to achieve something and the appearance of concrete evidence of success. Immature students are therefore easily misled into believing that they have truly mastered a craft of immense power and of great importance when, in fact, they have learned only its rudiments and nothing substantive at all. A student's quick climb from a state of complete ignorance about computers to what appears to be a mastery of programming, but is in reality only a very minor plateau, may leave him with a euphoric sense of achievement and a conviction that he has discovered his true calling. The teacher, of course, also tends to feel rewarded by such students' obvious enthusiasm, and therefore to encourage it, perhaps unconsciously and against his better judgment. But for the student this may well be a trap. He may so thoroughly commit himself to what he naively perceives to be computer science, that is, to the mere polishing of his

programming skills, that he may effectively preclude studying anything substantive.

Unfortunately, many universities have "computer science" programs at the undergraduate level that permit and even encourage students to take this course. When such students have completed their studies, they are rather like people who have somehow become eloquent in some foreign language, but who, when they attempt to write something in that language, find they have literally nothing of their own to say.

The lesson in this is that, although the learning of a craft is important, it cannot be everything.

The function of a university cannot be to simply offer prospective students a catalogue of "skills" from which to choose. For, were that its function, then the university would have to assume that the students who come to it have already become whatever it is they are to become. The university would then be quite correct in seeing the student as a sort of market basket, to be filled with goods from among the university's intellectual inventory. It would be correct, in other words, in seeing the student as an object very much like a computer whose storage banks are forever hungry for more "data." But surely that cannot be a proper characterization of what a university is or ought to be all about. Surely the university should look upon each of its citizens, students and faculty alike, first of all as human beings in search of—what else to call it?—truth, and hence in search of themselves. Something should constantly be happening to every citizen of the university; each should leave its halls having become someone other than he who entered in the morning. The mere teaching of craft cannot fulfill this high function of the university.

Just because so much of a computer-science curriculum is concerned with the craft of computation, it is perhaps easy for the teacher of computer science to fall into the habit of merely training. But, were he to do that, he would surely diminish himself and his profession. He would also detach himself from the rest of the intellectual and moral life of the university. The university should hold, before each of its citizens, and before the world at large as well, a vision of what it is possible for a man or a woman to become. It does

this by giving ever-fresh life to the ideas of men and women who, by virtue of their own achievements, have contributed to the house we live in. And it does this, for better or for worse, by means of the example each of the university's citizens is for every other. The teacher of computer science, no more nor less than any other faculty member, is in effect constantly inviting his students to become what he himself is. If he views himself as a mere trainer, as a mere applier of "methods" for achieving ends determined by others, then he does his students two disservices. First, he invites them to become less than fully autonomous persons. He invites them to become mere followers of other people's orders, and finally no better than the machines that might someday replace them in that function. Second, he robs them of the glimpse of the ideas that alone purchase for computer science a place in the university's curriculum at all. And in doing that, he blinds them to the examples that computer scientists as creative human beings might have provided for them, hence of their very best chance to become truly good computer scientists themselves.[8]

Finally, the teacher of computer science is himself subject to the enormous temptation to be arrogant because his knowledge is somehow "harder" than that of his humanist colleagues. But the hardness of the knowledge available to him is of no advantage at all. His knowledge is merely less ambiguous and therefore, like his computer languages, less expressive of reality. The humanities particularly

> "have a greater familarity with an ambiguous, intractable, sometimes unreachable [moral] world that won't reduce itself to any correspondence with the symbols by means of which one might try to measure it. There is a world that stands apart from all efforts of historians to reduce [it] to the laws of history, a world which defies all efforts of artists to understand its basic laws of beauty. [Man's] practice should involve itself with softer than scientific knowledge. . . . that is not a retreat but an advance."[9]

The teacher of computer science must have the courage to resist the temptation to arrogance and to teach, again mainly by his own example, the validity and the legitimacy of softer knowledge. Why

courage in this connection? For two reasons. The first and least important is that the more he succeeds in so teaching, the more he risks the censure of colleagues who, with less courage than his own, have succumbed to the simplistic worldviews inherent in granting imperial rights to science. The second is that, if he is to teach these things by his own example, he must have the courage to acknowledge, in Jerome Bruner's words, the products of his subjectivity.

Earlier I likened the unconscious to a turbulent sea, and the border dividing the conscious, logical mind from the unconscious to a stormy coastline. That analogy is useful here too. For the courage required to explore a dangerous coast is like the courage one must muster in order to probe one's unconscious, to take into one's heart and mind what it washes up on the shore of consciousness, and to examine it in spite of one's fears. For the unconscious washes up not only the material of creativity, not only pearls that need only be polished before being strung into structures of which one may then proudly speak, but also the darkest truths about one's self. These too must be examined, understood, and somehow incorporated into one's life.

If the teacher, if anyone, is to be an example of a whole person to others, he must first strive to be a whole person. Without the courage to confront one's inner as well as one's outer worlds, such wholeness is impossible to achieve. Instrumental reason alone cannot lead to it. And there precisely is a crucial difference between man and machine: Man, in order to become whole, must be forever an explorer of both his inner and his outer realities. His life is full of risks, but risks he has the courage to accept, because, like the explorer, he learns to trust his own capacities to endure, to overcome. What could it mean to speak of risk, courage, trust, endurance, and overcoming when one speaks of machines?

NOTES

Notes to Introduction

1. M. Polanyi, *The Tacit Dimension* (New York: Doubleday, Anchor ed., 1967), pp. 3-4.
2. This "conversation" is extracted from J. Weizenbaum, "**ELIZA**—A Computer Program For the Study of Natural Language Communication Between Man and Machine," *Communications of the Association for Computing Machinery*, vol. 9, no. 1 (January 1965), pp. 36-45.
3. K. M. Colby, J. B. Watt, and J. P. Gilbert, "A Computer Method of Psychotherapy: Preliminary Communication," *The Journal of Nervous and Mental Disease*, vol. 142, no. 2 (1966), pp. 148-152.
4. *Ibid.*
5. T. Winograd, "Procedures As A Representation For Data In A Computer Program For Understanding Natural Language." Ph.D. dissertation submitted to the Dept. of Mathematics (M.I.T.), August 24, 1970.
6. J. Weizenbaum, 1972.
7. Hubert L. Dreyfus, *What Computers Can't Do* (Harper and Row, 1972).
8. Hannah Arendt, *Crises of the Republic* (Harcourt Brace Jovanovich, Harvest edition, 1972), pp. 11 *et seq.*