24. [Introduction]
From *Computer Power and Human Reason*
From Judgment to Calculation

◊03 49

◊04 65

"Technology Assailed by Angry Humanist" would be an apt headline in *The Onion*, the satirical weekly newspaper known to most of its readers through the Web. Dire warnings about computers have frequently issued from non-computer-using authors who have only superficially considered, or even experienced, new media. There is some humanistic commentary on technology, such as that of Langdon Winner (◊40), that is more informed. The discussion of new media pitfalls offered by Joseph Weizenbaum comes from someone with a different sort of background, however. Weizenbaum did not write as an observer and commentator outside of computing. He programmed the most famous chatterbot in the history of computing and then perceived dangerous uses of the system he himself engineered. He followed Norbert Wiener (◊04), another MIT professor, in demanding that scientists and technologists take responsibility for the use of that which they discover and develop.

Human languages did not make much of a place for themselves in computing during the 1950s. It wasn't until the early 1960 that word processing began to take shape; in 1963, for instance, an early program for writing on the computer was developed by hackers at MIT. Although some question-answering systems had been programmed earlier, as Weizenbaum documents in his book, the first more general conversational computer program of the sort that Alan Turing envisioned (◊03) was the one Weizenbaum created from 1964 to 1966. This system, called Eliza, ran a set of scripts called *Doctor* ⊗ and impersonated a psychotherapist, becoming notorious and leading Weizenbaum to profoundly reassess his ideas about computing. *Doctor* wasn't foolproof, but that set of instructions allowed Eliza to plausibly carry on some conversations, posing as a noncommittal Rogerian who would ask the user to reflect on whatever comments were offered. While Turing's guess that a thinking machine would be around in 2000 may not have been right on target, within more limited contexts, Turing's prediction that computers would plausibly interact with people using language as an interface was borne out long ago, by Weizenbaum's work in the 1960s.

367

Eliza and the Doctor script are included in a Java implementation on the CD.

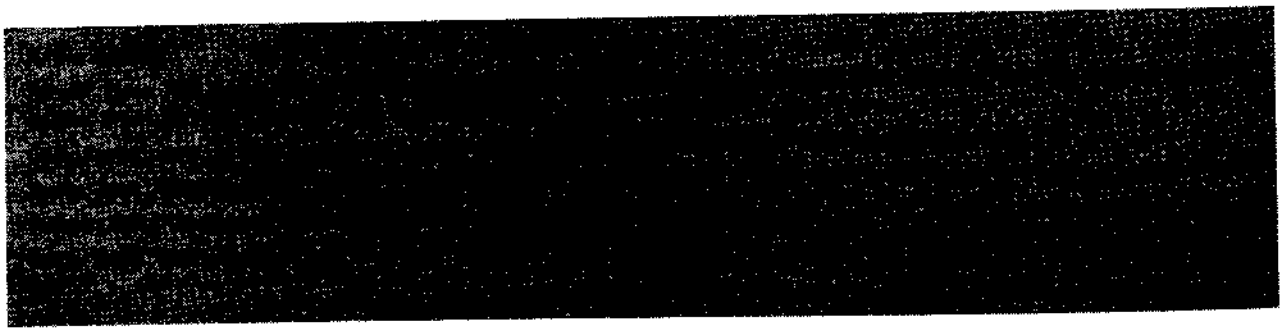
◊34 499

The concern that machines will take over not just the jobs that provide us income, but also those cognitive and emotional functions we closely associate with humanity, is a particular worry of the computer era—one that was highlighted for Weizenbaum by the way some suggested that *Doctor* should be employed as an actual therapist. Neil Postman explores what Weizenbaum calls the "ever more mechanistic image" of humanity in his book *Technopoly*, in which he notes that the influence of computing is seen not only in examining specific hardware and software technologies but in the way we employ language that has been appropriated for use in computing. Even as they bring material benefits, new media and other computing innovations shift concepts like "belief" and "virus" so that they are more associated with digital computing and less connected to their original humanistic or biological meanings. Weizenbaum's specific concern is that people who see the computer as able to assume the intimate and human role of the psychotherapist are unable to draw the boundaries between the proper use of computer technology and "computer applications that either ought not to be undertaken at all, or, if they are contemplated, should be approached with the utmost caution" (268).

◊40 587

Sherry Turkle (◊34) gives a psychoanalytic perspective on Eliza in her *Life on the Screen*, reaching a different conclusion. Kenneth Colby's view that computer therapy could be useful, after all, was not a snap judgment. Colby was a computer scientist and psychiatrist; he later used Eliza to create a simulated schizophrenic, *Perry*, to aid in psychiatric study, and created a therapeutic conversational system *Depression 2.0*, released in 1992. Although this is not mentioned in *Computer Power and Human Reason*, Turkle notes that Colby was Weizenbaum's original collaborator on Eliza (106). Turkle found

in her research that people who appreciated Eliza considered it as "a kind of diary or mirror" (108) and did not end up with the distorted view of humanity that Weizenbaum feared—but in her book, she also considers how human-computer interaction does change our concepts of what is alive, what is human, and what we ourselves are. Janet Murray's reevaluation of *Eliza/Doctor* as literary art offers another way that Weizenbaum's accomplishment can be recognized as opening up possibilities for creative expression rather than as disrupting our notions of humanity.
—NM



Original Publication
Introduction. *Computer Power and Human Reason: From Judgment to Calculation*, 1–16. San Francisco: W.H. Freeman and Company, 1976.

From *Computer Power and Human Reason*

From Judgment to Calculation

Joseph Weizenbaum

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*]

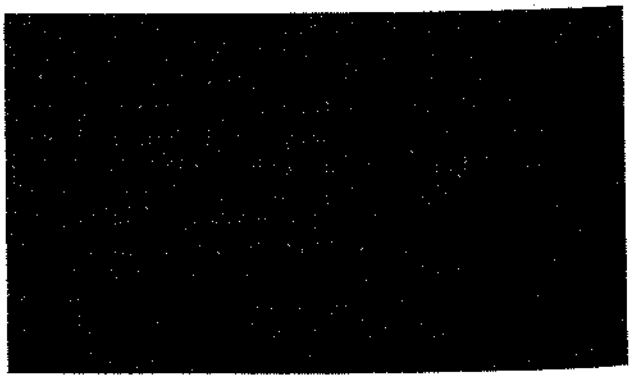
I had thought it essential, as a prerequisite to the very possibility that one person might help another learn to cope with his emotional problems, that the helper himself participate in the other's experience of those problems and, in large part by way of his own empathic recognition of them, himself 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 clockwork. 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 magician), 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.[1] 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.

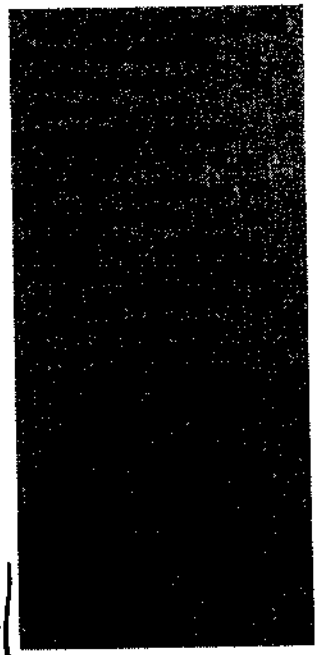
It may be that human values are illusory, as indeed B. F. Skinner argues. If they are, then it is presumably up to science to demonstrate that fact, as indeed Skinner (as scientist) attempts to do. But then science must itself be an illusory system. For the only certain knowledge science can give us 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.

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.

Notes

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.*