Partisans and Critics of a New Science: The Case of Artificial Intelligence and Some Historical Parallels

Everywhere in our society—at the supermarket, in the stock market, in mathematics courses—we see the presence of the computer. We are told that we are entering a new age, that of the Computer Revolution, and that the field known as Artificial Intelligence is at the forefront of that revolution. Its practitioners have proclaimed that they will solve the age-old problems of the nature of human thought and of using technology to build a peaceful and prosperous world. These claims have provoked considerable controversy.

The field known as AI is, of course, vast, but the subject matters that have fueled the most debate are these: computer systems that exhibit behaviors that, if exhibited by people, should be called "intelligent"; and "expert" computer systems that, though not exhibiting a broad range of such behaviors, can, because of speed and size of memory and built-in knowledge, outperform humans at specific tasks. The chief implications of research in these areas, according to its advocates, and the claims most viewed with alarm by the research's critics are that AI has a new way of understanding the nature of human beings and their place in the universe and that the "technology" of task-oriented AI—sometimes called *expert systems*, sometimes *knowledge engineering*—will be able to solve the outstanding problems facing man and society.

Critics of AI enthusiasts perceive these claims as wrong: as based, first, on assumptions about the nature of man and society that are inadequate at best and dangerously misleading at worst; and, second, as overly optimistic about the feasibility of finding technical solutions to human problems. In the first case, the critics say, advocates of AI reify a mechanical model of man and of society, one that is not only insufficient but that encourages the treatment of individuals mechanically and as parts of a larger machine. In the second case, they propose technical solutions to

what may not be technical problems, and they underestimate the costs of technical innovations while overestimating the benefits.

Upon listening to a debate like this, a historian might well ask whether the claims of the partisans of the science—and of their critics—are solely reactions to the current and probable future state of their science, or whether they are drawing instead on the rhetoric and ideas of other debates about somewhat different topics. If the latter, how did those earlier debates—on which more of the returns are in—actually come out?

In this paper, I hope to provide some historical perspective on the implications of Artificial Intelligence by describing a pattern followed by similar debates about the implications of scientific breakthroughs earlier in history. I will present several examples to show the frequency with which this pattern has been followed. But the purpose of giving these examples is not only to demonstrate that this pattern has occurred in the past. I want also to suggest that causes similar to those that operated in the past are working now; that some of the content of past debates has helped shape the current one; and, finally, that recognizing the pattern will help clarify what is really at issue now. In short, I hope to illuminate the present debate over AI by showing that it shares many characteristics with past debates between partisans of new scientific approaches and critics of those approaches.

The pattern followed is roughly this: (1) The new methods solve some outstanding set of problems despaired over in the past, and these successes are impressive to outsiders and inspiring to those who have achieved them. (I call this first stage *technical success*.) (2) The practitioners extrapolate their successes to other fields (they have found the language of the book of nature, or the secret of life); at the very least, they argue that those who use their methods in other areas will succeed by these new means where earlier investigators have failed. (I call this second stage *extrapolation*.) (3) In reaction, antiscientific currents that always surround the sciences— currents hostile even to successes of the delimited kind first mentioned— are intensified by the extreme claims made in the second stage. (4) Finally, more serious critics enter the fray—critics often from within the scientific community itself, who perceive unfounded enthusiastic claims, claims whose consequences—with no apparent scientific warrant—threaten cherished values.

There are three other striking phenomena. First, the promises of the first stage—successes in the original area and its nearby neighbors—are

in fact fulfilled, but the claims of the second stage largely are not; and there are costs even of the first. Second, the mature science growing out of the initial success consolidates the earlier gains and continues. Third, the enthusiasm and inspiration of the second stage seem valuable in consolidating the gains of the first stage of initial success, even though their wide-ranging predictions are not fulfilled.

The debates also make reference to historical analogies taken from similar debates in the past, and in so doing, they often blur the distinction between the stage of initial success and the second, more enthusiastic stage. Critics opposing the second-stage extrapolation are often accused of opposing the first, and are called Luddites or persecutors of Galileo. Keeping the distinctions between the four stages clear, even though there may be some blurring near the boundaries, would improve the current debates over the support that society should give to fifth-generation AI research, the human tasks that should be entrusted to computers, and the information that artificial intelligence might give us about the nature of man. It is important that these debates not be carried out with arguments like "They laughed at Fulton" on the one hand and "AI is the new alchemy" on the other. Of course, history can help, but one must first understand that history.

To support these general points, I shall outline a set of examples from the history of science so that we can then describe recent debates over AI with this historical perspective. I will in particular discuss the methodological revolution in seventeenth-century science; the eighteenth-century "spirit of systems" and the visions of society stemming from it; the Industrial Revolution; and Darwinian evolution and the extrapolations—social Darwinism, racism, atheistic materialism—from it. Finally, I will take up the key ideas at the early stages of AI in the work of Alan Turing; the claims of AI pioneers like Herbert Simon, Alan Newell, John McCarthy, and Marvin Minsky, and some of their successors; some reactions to the overly enthusiastic claims; and, finally, the judicious, historically informed criticism of AI by computer scientist Joseph Weizenbaum.

The first example is to be found in the period of the scientific revolution—in particular, the methodological revolution of seventeenth-century science. Let us begin by measuring the first stage: its successes in its own sphere. As recently as the sixteenth century, Michel de Montaigne could speak of the futility of learning and could use the Copernican system and the fact that it had ancient predecessors as an example

of the way ideas about nature came and went as did other fashions (*Defense of Raymond Sebond*, 86-87). According to Montaigne, the sciences had established no real knowledge. But to Francis Bacon, this lack of past success called not for despair but for a new approach: "Things which have never yet been done can be done [only] by methods which have never yet been tried" (*Novum Organum*, aphorism vi). And, indeed, the mechanical philosophy did things that had never been done. Using the Cartesian method of "analysis"—studying macroscopic phenomena by resolving them into their component parts of bits of matter in motion—men like Pascal, Torricelli, and Boyle were able to explain, by mechanical means, a whole range of phenomena previously requiring notions like the abhorrence of a vacuum (Westfall 1971, 44-49). William Harvey, treating the heart as if it were a pump and the blood as though it were an ordinary fluid, demonstrated the circulation of the blood. Descartes saw Harvey's work—though Harvey himself was far from being a mechanist—as a triumph of the mechanical and analytic methods (Westfall 1971, 93; *Discourse on Method*, 47-55). Cartesians saw the method as opening all of nature to being reduced to Descartes's principles, with Descartes himself believing that he had exhibited the mechanism of the solar system. Some of his followers took the mechanical philosophy even further: LaMettrie in 1747 spoke of man as a machine, and Hobbes (unlike Descartes) saw even thought as a purely mechanical process (*Leviathan*; in Burtt 1939, 143-44).

But the outstanding predictive success of the mechanical philosophy when applied to the vacuum or (with Borrelli) to the mechanical advantage of muscles did not extend to psychology, or even to all of physics. Descartes's prediction that a science of nature on his principles was almost with his reach (*Discourse on Method*, 68) was an example of extrapolation beyond the domain of success. So were the views that man was wholly mechanical, and so were the views that the Cartesian principles sufficed to explain the phenomena of the solar system. As is well known, reactions even against the initial stage of predictive successes of science occurred in the seventeenth century, from the fabled few who refused to look through Galileo's telescope to the poets who grieved that "new philosophy calls all in doubt" (Donne, "Anatomy of the World"; quoted by Koyré 1957, 29). But there were other opponents of the new philosophy who appreciated what the scientists had accomplished but who saw no reason to believe that it extended as far as the scientists claimed. Blaise

Pascal made clear that, though scientific reasoning had its place, there was much in the situation of man in the world, and man's understanding of himself, that was not amenable to reason. Pascal's *esprit de finesse* recognized intuition and tacit ways of knowing as just as valid in their sphere as mathematical reasoning is in its proper sphere (Pensées 1).

As for Newton and the Newtonians, they found the Cartesian insistence that causes must be mechanical as the same sort of excess. Perhaps the cause of gravity was not known; this did *not* mean that it was reducible to mechanism, any more than the fact that a clock *could* be run by a spring meant that a *given* clock was not actually moved by a weight. (For this Turing test for pendulum clocks, see Roger Cotes's Preface to Newton's *Principia* [pp. xxvii-xxviii].) Besides, as Newton had shown, Cartesian mechanism had not been able to account for all the phenomena of the solar system (notably Kepler's laws), whereas gravity seemed really to exist, though its cause was unknown, and gravity could explain the motions of the heavenly bodies, the fall of terrestrial objects, and the interaction between them in the movement of the tides (Newton's *Principia*, 547). Thus the methodological imperialism of the Cartesians had not in fact produced the promised understanding of the universe, much less of man. That critics like Pascal and the Newtonians waged part of their battle on behalf of their views of religion does not invalidate their conclusions about the limited scope of the successes of the mechanical philosophy.

My second example is found in the "spirit of systems" characteristic of such thinkers of the Enlightenment as Voltaire, Condorcet, Comte, and St.-Simon. Buoyed up by the successful application of reason in the realm of nature, and by the analysis of macroscopic phenomena in terms of simple statements about the elements, they hoped to achieve in the same way a science of society capable of remaking the world in a rational way. As one opponent, A. Thibaudet, put it, they "ventured to create a religion as one learns at the Ecole [polytechnique] to build a bridge or a road" (Hayek 1955, 113). They had, however, left out of their calculations the passions unleashed by, for instance, the French Revolution, and the tenacious holding power of old institutions and ideas and the powerful people who wanted those preserved. In particular, Voltaire's faith in the power of reasonable arguments to persuade anybody who heard them, based on the success of Newtonian science (Voltaire 1764, article "Sect"), was not supported by events. The institutions of the past, as Burke pointed out, and the forces preserving them, as Marx pointed out, had too much

staying power. The philosophes, to be sure, left a lasting legacy of new political and social institutions, but things did not work out as they had predicted and hoped. Representing the scientists of our fourth stage, A.-L. Cauchy in 1821 said to his colleagues: "Let us enthusiastically cultivate the mathematical sciences, without trying to extend them beyond their domain; and let us not imagine that one can attack history with formulas, nor give as sanctions to morality theorems of algebra or of integral calculus" (*Cours d'analyse*, vii).

Another example is to be found in the Industrial Revolution. Indeed, the parallel between the Industrial Revolution and the "Information Revolution" we are now supposedly experiencing is made repeatedly in the AI community. The Industrial Revolution in Britain, whose history has been thoroughly written (see, e.g., Deane 1979), was billed as arising from science (though there was more rationalism of the Cartesian variety and analysis of systems using the concept of "division of labor" than much new science, at least before the 1850s). Nineteenth-century partisans of the Industrial Revolution, drawing on Baconian prophecies of the benevolence of applied science and science's ability to improve the lot of all mankind, theorized far beyond their data. Industrialization was certainly bringing about new jobs and providing new products and services, which on balance raised the average standard of living (Deane 1979, 272-95). But it also altered people's lives, took employment from some workers, and produced a new type of job and a new relation between employer and employee. There was incredible urban filth as people were crowded together without adequate water and sewers (Deane 1979, 260-61), and the worker worked at the rhythm of the factory, which led to feelings of alienation. Charles Dickens reflects these widespread feelings in his novel *Hard Times*, when the worker Stephen Blackpool, whose trade union expelled him because he wanted to go his own way and who later died after falling into an old mineshaft that it had not been cost-effective to make safe, expressed his view of the inhumanity of the system: It is all a "muddle." And, Adam Smith to the contrary, it will not fix itself: "Nor yet lettin alone will never do 't. . . . Rating 'em [workers] as so much Power, and reg'latin 'em as if they was figures in a soom [sum], or machines, wi'out loves and likens, wi'out memories and inclinations, wi'out souls to weary and souls to hope . . . this will never do 't" (pp. 115-16). The enthusiasts of the Industrial Revolution consistently underestimated these costs; rather than working to minimize the effect of the costs, they often

justified them as part of the workings of the system of efficient organization of labor, whose ultimate tendency, they thought, was better (at least statistically) for everyone.

These tendencies continued into the twentieth century, notably with the work of F. W. Taylor on what is called "scientific management." The organization of production into separate tasks, each one optimized, is the key to efficiency, said Taylor in 1911: "In the past the man has been first; in the future the system must be first" (Pursell 1969, 246). Listen to Taylor instructing a workman to listen to the scientific manager: "A high-priced man has to do exactly as he's told from morning till night.... You will do exactly as this man tells you.... When he tells you to pick up a [piece of pig iron] and walk, you pick it up and you walk, and when he tells you to sit down and rest, you sit down.... And what's more, no back talk" (Pursell 1969, 246). This is *not* Dickensian satire.

Even enemies of Taylorism often conceded that, like it or not, it was "the most efficient means of production" (Keir 1917; in Pursell 1969, 255). However, the neglect of the human costs of doing a mechanized, though optimized, set of tasks surely is one of the reasons workers feel alienated, and many labor historians hold this feeling to be at the root of inefficiencies ranging from goldbricking to labor strife. The famous Hawthorne experiments of the late 1920s, in fact, showed that workers worked better when management paid more attention to them (Noble 1977, 319). This research gave rise to a set of new disciplines, like industrial psychology, and showed that Taylorism, even from the standpoint of maximizing worker efficiency, was inadequate. Similarly, it has recently been suggested that cooperative models, as used in Japan, are more efficient than the manufacturer-initiated, top-down organization characteristic of nineteenth-century European industry. One more point needs to be made to underline the applicability of our pattern to this example: it is true that some critics of industrialism wanted to reverse the whole process, but others, like George Orwell—who certainly had no love for mechanization (Orwell 1937, chap. 12, esp. p 160)—stressed that technological advances had now made it at least possible to raise everyone's standard of living "if we chose to set our minds to it" (Orwell 1943, 213) and criticized industrialism principally for its failure to provide the predicted benefits to all.

My last example is Darwinism. Charles Darwin and Alfred Russel Wallace proposed the idea of evolution by natural selection to explain the origin of biological species, their geographic and fossil distribution, and

their adaptation to the conditions of life. Some of their followers used the removal of God from the creation of species to advocate a materialistic biology from which all nonmechanical causes were to be banished. T. H. Huxley went so far as to predict that, just as there was a mechanical equivalent of heat, so we would find a "mechanical equivalent of consciousness" (Huxley 1870, 191). Others used evolution to argue that the survivors in economic competition were the "fittest," or to justify the racial superiority of Caucasians, which, after Darwin, could no longer be scientifically justified by the special creation of the white race (Stanton 1960; Hofstadter 1955). The religious attacks on evolution were strongly reinforced because of these extensions—with ominous consequences—of the doctrine, as the examples of the anti-slavery Bishop Wilberforce and the populist and anti-imperialist William Jennings Bryan show. The eugenics movement in the United States provides another example of the unwarranted extension of evolutionary ideas. During the debate in Congress over immigration restriction in 1924, the eugenicists made the selective restriction against southern and eastern European immigration "seem like a biological imperative" (Ludmerer 1972, 113).

Critical responses from within the scientific community to these extrapolations of evolution ultimately were to prevail. Thus Asa Gray in the nineteenth century (Dupree 1959, 375-77) showed that an evolutionist could be a Christian and not a reductionist, saying that what "science removes from the supernatural to the natural [is not therefore] lost to theism." Similarly, Kirtley Mather in the twentieth century, testifying at the Scopes trial as a geologist and a prominent Baptist layman, showed by example as well as precept that one could be a professing Christian as well as an evolutionist (Ginger 1969). Again, the geneticist Herbert Spencer Jennings (the only scientist to testify against immigration restriction before Congress in 1924) and the anthropologist Arthur Kroeber argued in the 1920s that the eugenicists' arguments went far beyond anything justified by genetic science (Ludmerer 1972, 113; Cravens 1978, 105). By the 1930s, most American geneticists had repudiated eugenics, attributing to it, in the words of geneticist Edward M. East, "too great a tendency to push ahead of the facts" (Ludmerer 1972, 123).

The examples I have cited could be multiplied many times, from the Pythagorean idea that all of nature could be reduced to simple mathematics to the Skinnerian idea of learning by reinforcement and the resulting predictions of the imminent successful introduction of teaching machines.

(See also Lorraine Daston's paper on the probability of judgments in the present volume.) The specific examples I have chosen to discuss are chosen not only for their importance, but because they have a direct bearing on the current debates over AI. Not only is the structure of the debates similar, but the same themes—rationality and mechanism as a complete picture of reality, technological development as the solution to social problems—have been revived and redeveloped in the present debates. Let us now look at the recent history of Artificial Intelligence: at some successes, at some claims made, and at the types of critical responses that have been produced.

Borrowing the idea of a completely mechanized thought process with no room for free will from the mechanical world views of the seventeenth and eighteenth centuries, and drawing also on logic and on his own formal analysis of computability, Alan Turing saw no reason to conclude that a machine could not think, if by *think* one meant to imitate a thinking human being to the point where the machine could fool a human conversational partner into believing it to be another human (the Turing test; Turing 1956). The mathematical abstraction he had created, the Turing machine, enables one to prove many results about what can and cannot be computed.

> Given the initial state of the machine and the input signals, it is always possible to predict all future states. This is reminiscent of Laplace's view that from the complete state of the universe at one moment of time ... it should be possible to predict all future states. The prediction we are considering is, however, rather nearer to practicability than that considered by Laplace. (Turing 1956, 2105)

Turing suggested that a computer with 10^9 memory locations would in about 1990 have a 70 percent chance of passing the Turing test (2107) and thus a 70 percent chance of being able to think. Responding to the possible objection that thinking requires a soul, Turing said that "we should not be irreverently [usurping] God's power of creating souls, any more than we are in the procreation of children; rather we are, in either case, instruments of His will providing mansions for the souls that he creates" (2108). But this is all beside the point: "I am not very impressed with theological arguments whatever they may be used to support," he says, dismissing them all with a reference to the case of Galileo.

The mind, then, seems for Turing to be a Turing machine. The adult mind might be hard to imitate, Turing continues, since the processes that

have brought it to the "state that it is in" include "the initial state of the mind, say at birth," education, and other experiences. Education and experience are not provided in the most efficient possible manner; this could be done more rationally if we started with the initial state: "Presumably the child-brain [has] so little mechanism that something like it can be easily programmed" (2119).

Turing's views on computability and thought as both being formally describable, rule-based processes appeared to some philosophers, notably J. R. Lucas, to be similar to the old mechanistic ideas, though in a new dress. Even if these new mechanists do not assert that minds come out of matter, said Lucas, they do claim that minds "operate according to 'mechanical principles'"—that is, think that the whole is just the sum of the operations of the separate parts and that the operation of each part is determined by previous states or by random choices between states (Lucas 1961, 126). I have discussed this controversy elsewhere (Grabiner 1984, 1986). For our present purposes, the key point is that Lucas's opposition to Turing's ideas came from Lucas's desire to refute both "traditional materialism and its modern mechanist version" (Lucas 1968, 156), a "mechanism" that is not implied by, but extrapolated from, Turing's research in computability theory.

Beginning in 1956 with the Dartmouth conference that launched AI as an organized discipline in the United States many more arguments characteristic of the "extrapolation" stage began to appear. The original proposal to the Rockefeller Foundation for the conference said: "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it" (McCorduck 1979, 93). The key ideas used to "precisely describe" intelligence are that the human mind is a system, either like a machine or like Laplace's world-system, hierarchically organized and programmed by its environment; thought takes place by the rule-generated manipulation of formal symbols. The mind's organization, like the Tables of Discovery of Francis Bacon (*Novum Organum*, aphorism cii), provides for empty slots into which pieces of information from the world (disentangled from their context) can be placed. For instance, John McCarthy: "One of the basic entities in our theory is the situation. Intuitively, a situation is the complete state of affairs at some instant in time. The laws of motion of a system determine all future situations from a given situation" (McCorduck 1979, 213).

Similarly, Simon and Newell state, "The free behavior of a reasonably intelligent human can be understood as the product of a complex but finite and determinate set of laws" (quoted by Dreyfus 1979, 170). And, in an often-quoted passage: "A man, 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.... I myself believe that the hypothesis holds *even for the whole man*" (Simon 1981, 65; my emphasis).

We should acknowledge the initial success of early AI programs in imitating some aspects of human thought: solving mathematical and logical problems; "memorizing" nonsense syllables in ways similar to those psychologists find in human subjects; conducting dialogues about limited worlds; answering questions about brief, formula-based stories; and playing checkers and chess. All this was accompanied by enthusiastic predictions. Herbert Simon in 1960 proclaimed that in ten years a computer would be world chess champion, discover and prove an important new mathematical theorem, and translate natural languages, and also that most psychological theories would be embodied in computer models (quoted by Dreyfus 1979, 81-82). Marvin Minsky proposed that computers would write good music and draw significant pictures as soon as aesthetic analysts do something about "the traditionally disgraceful state of analytic criticism in the arts" (Minsky 1968, 12; quoted by Weizenbaum 1976, 157). In words deliberately (and only too successfully) provocative, John McCarthy asked, "What do judges know that we cannot eventually tell a computer?" (McCorduck 1979, 324), and Minsky called the brain "a meat machine" (McCorduck 1979, 170).

Let us now turn to some of the prominent critics. An attack was made on the entire AI program by Hubert Dreyfus, who claimed that the "successes" were trivial and that predictions about uses in the future would never be fulfilled. His weapons were chiefly Wittgensteinian language theory and a phenomenological view of the world. John Searle attacked AI enthusiasts on other grounds: he argued that even the successful passing of the Turing test by a formal system—that is, a formal system capable of making people believe it was thinking—would not mean that the machine really understood anything. He asks us to imagine a human being "simulating" natural-language understanding by simply matching symbols according to a book of rules, without ever knowing what the symbols mean. The person would still not understand. This rejection of

behavioral equivalence on Searle's part is not an attack on the actual accomplishments of AI research, but on the interpretation the researchers place on that work (Searle 1980; Grabiner 1984, 474-76).

The work of MIT computer scientist Joseph Weizenbaum itself produced a capsule version of our pattern. Weizenbaum in 1966 had published a paper on a program called ELIZA that illustrates its ability to match patterns in English sentences by mimicking the "mirroring of statements" of a therapist of the school of Carl Rogers (Weizenbaum 1966). Others, not Weizenbaum, did the enthusiastic extrapolation of his work: that it had passed the Turing test, that it meant that automated psychiatry was just around the corner (Grabiner 1984, 1986; Weizenbaum 1976, 180). Weizenbaum was appalled. He has since devoted much of his professional life to warning against extrapolating what he calls the "computer metaphor" to areas of human life to which, he argued, it does not pertain. Objecting to the contention that the mind is just a computer operating on formal symbols, Weizenbaum argued instead that humans use language always in context—notably "in a context of experiences like love and trust that machines cannot share" (Weizenbaum 1976, 208-9). And Weizenbaum, being himself involved in the field, made a criticism of the basis for AI's extrapolations that the philosophers cited earlier had been unable to make: the most successful AI programs, he said, lacked strong theory-based models of human intelligence, being instead largely collections of ad hoc programming tricks (Weizenbaum 1976, 232); therefore, such programs, however well they seemed to imitate certain human behaviors, did not in any sense "explain" the nature of human thought. I will return to Weizenbaum's critique of AI later on.

As AI has developed, what some people have characterized as a paradigm change has come about (Duda and Shortliffe 1983). There has, in particular, been a realization that building programs that can learn the way a child does and that can proceed with common sense in everyday situations is very hard indeed. However, programs embodying human expertise in specific areas, with the knowledge organized according to AI-based theories of knowledge representation and using the AI techniques of heuristic search, have been developed that perform impressively. DENDRAL in chemical structure, MYCIN in diagnosing bacterial infections, and PROSPECTOR in mineral exploration are well-known examples (Duda and Shortliffe 1983; Hayes-Roth et al. 1983; Feigenbaum and McCorduck 1983). These examples of applied AI, which are called expert systems or

knowledge engineering, somewhat resemble the way the manufacturers of the Industrial Revolution applied some physical principles and the idea of systems; but the amount of science involved here is considerably greater. Many people working in expert systems have drawn back from the claims made earlier for AI and have conceded that whether computers can imitate the full range of human intelligent behavior is "unknown and perhaps . . . forever unknowable" (Albus 1981, 299; cf. Hayes-Roth et al. 1983, 3).

But this modesty of theoretical goals has not precluded another set of enthusiastic predictions, this time reminding us of the prophets of the Industrial Revolution: that expert systems provide an "unprecedented opportunity" (the subtitle of the recent book *Intelligent Systems* (1983) by J. E. Hayes and Donald Michie) to solve the world's problems. Thus intelligent computers will revolutionize education (Papert 1980; Schanck 1984), will solve world problems like terrorism as they read and analyze news stories (Schanck 1984, 220). Edward Feigenbaum and Pamela McCorduck say that, like Adam Smith, they "also write inspired by a machine . . . . An intelligent machine [is] the engine that will produce the new wealth of nations" (1983, 17). The Japanese futurist Yoneji Masuda predicts that technical advances in computing mean that "as the 21st century approaches . . . the possibilities of a universally opulent society being realized have appeared in the sense that Smith envisioned it . . . a society that actually moves toward a universal society of plenty" (Masuda 1985, 625). Thus the claims about what computer technology will make possible are expanded past the actual capabilities of today's computers, and the nontechnological difficulties that will certainly arise are minimized. In an odd use of history, potential ill effects are dealt with by detaching them from computers. Thus it is claimed that there has already been considerable unemployment and dehumanization without computers (e.g., Schanck 1984, 235), so there is no reason to fear that computers will make things worse, and they might help. This follows the same pattern that we saw in the Industrial Revolution, in the Enlightenment's desire to build a commonwealth based on reason, and in Francis Bacon's view that the fruits of science could only help, not hurt mankind (*Novum Organum* aphorism cxxix). It is hard, after Auschwitz and Hiroshima, to accept this kind of optimism in exactly this form; there are, at the very least, costs that will need to be considered in assessing the proposed benefits.

The sociologist Sherry Turkle has characterized the "extrapolation"

stage of AI research in words that equally describe the mechanical philo-sophers of the seventeenth century, the social theorists of the age of the French Revolution and of Taylorism, and the apostles of social Darwinism and eugenics—in terms of the concept of a *colonizing discipline*. "Being in a colonizing discipline," she writes, "first demands and then encourages an attitude [of] intellectual hubris . . . , intellectual principles . . . univer-sal enough to give you the feeling that you have something to say about everything . . . [and] confidence that your knowledge makes the 'traditional wisdom' unworthy of serious consideration." This "so fundamentally changes the rules of every game in the social and behavioral sciences that everything that came before is relegated to a period of intellectual imma-turity." And "you have to feel that nothing is beyond your intellectual reach if you are smart enough" (Turkle 1984, 251-52; cf. Descartes's *Discourse on Method*, 68).

Joseph Weizenbaum, as his continuing critique of AI and expert systems shows, is aware of the historical pattern I have been discussing. The in-troduction of computers into our society, says Weizenbaum, "merely rein-forced 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" (1976, 11). And this matters because of the way these views lead people to treat their fellows and to view themselves as programmed and thus not responsible for their actions (1976, chap. 10).

Responding to the methodological universalism of AI, Weizenbaum argues that what AI has done so far "is build a machine that performs certain specific tasks, just as, say, 17th-century artisans built machines that kept time," fired cannon balls, etc. They would have been mistaken, he continues, had they concluded that "because their machines worked, they had validated the idea that the laws of the universe are formalizable" in the same terms. But AI people, he concludes, advance "exactly these mistaken ideas about the machines [they have] succeeded in building" (1976, 196).

Weizenbaum would doubtless agree with Pascal with the *esprit géome-trique* has its proper realm: "Nothing is wrong with viewing man as an information processor," writes Weizenbaum, "providing, however, that we never act as though any single perspective can comprehend the whole man" (1976, 140). "Physicists, mathematicians, and philosophers" all ac-cept that "there are limits to how far the world can be comprehended" in what he calls Leibnizian terms, where knowing the initial conditions

lets us fully predict the future (1976, 221). But, as Aldous Huxley warned, "The popular Weltanschauung of our times contains a large element of what may be called 'nothing-but' thinking. Human beings . . . are nothing but bodies, animals, even machines . . . . Mental happenings are nothing but epiphenomena" (quoted by Weizenbaum 1976, 129). Max Horkheimer sums this all up to Weizenbaum's satisfaction: too many people believe that there is just one authority, namely science, and, furthermore, a science conceived merely as "the classification of facts and the calculation of probabilities" (Horkheimer 1947; quoted by Weizenbaum 1976, 252). Weizenbaum summarizes his own position: "The computer is a powerful new metaphor for helping us to understand many aspects of the world, but . . . it enslaves the mind that has no other metaphors and few other resources to call on." The computer scientist "must teach the limitations of his tools as well as their power" (1976, 277).

Before ending our discussion of Weizenbaum's ideas, let me remind the reader that there is nothing inevitable about the timing in the pattern I have described. In particular, the fourth-stage critics may be rare, and they do not necessarily win. As H. S. Jennings was unsuccessful in his lone opposition to immigration restriction in 1924, so Weizenbaum now is almost alone in the computer science community.

What, then, are we to conclude from the pattern common to all these examples from the seventeenth-century mechanists to AI? It is possible, as Max Weber has said, that scientific research must be a vocation and that the researcher must be passionately committed to the belief that the outcome of his next piece of research is the most important thing in the world. But as the rest of us try to understand the implications of artificial intelligence for our view of ourselves and for the future of our complex world, we should remember that, though the sciences have over time greatly enlarged the domain in which they can successfully predict and explain phenomena, the enthusiastic predictions about extending methods successful in one area to solve society's problems and to understand the world have generally not come to pass.

References

Albus, James. 1981. *Brains, Behavior, and Robotics.* Peterborough, N.H.: Byte Books.

Bacon, Francis. 1620. *Novum Organum.* Reprinted in (Burtt 1939).

Burtt, E., ed. 1939. *The English Philosophers from Bacon to Mill.* New York: Modern Library.

Cauchy, A.-L. 1821. *Cours d'analyse de l'école royale polytechnique.* Part 1. *Analyse algébrique.* In *Oeuvres complètes.* Paris: Gauthier-Villars, ser. 2, vol. 3, 1897.

Cravens, H. 1978. *The Triumph of Evolution*. Philadelphia: University of Pennsylvania Press.

Deane, Phyllis. 1979. *The First Industrial Revolution*, 2d ed. Cambridge: Cambridge University Press.

Descartes, Rene. [1637] 1956. *A Discourse on Method*. Reprinted (with page numbers from Adam-Tannery edition) New York: Bobbs-Merrill.

Dickens, Charles. [1854] 1966. *Hard Times*. Reprinted New York: W. W. Norton.

Dreyfus, H. 1979. *What Computers Can't Do: A Critique of Artificial Reason*, 2d ed. New York: Harper and Row.

Duda, R., and Shortliffe, E. 1983. Expert Systems Research. *Science* 220: 261-68.

Dupree, A. Hunter. 1959. *Asa Gray*. New York: Atheneum.

Feigenbaum, E., and McCorduck, P. 1983. *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. Reading, Mass.: Addison-Wesley.

Forester, Tom, ed. 1985. *The Information Technology Revolution*. Cambridge, Mass.: MIT Press.

Ginger, Ray. 1969. *Six Days or Forever*. New York: New American Library.

Grabiner, Judith. 1984. Artificial Intelligence: Debates about Its Use and Abuse. *Historia Mathematica* 11: 471-80.

———. 1986. Computers and the Nature of Man: A Historian's Perspective on Controversies about Artificial Intelligence. *Bulletin of the American Mathematical Society* 15: 113-26.

Hayek, F. 1955. The Source of the Scientistic Hubris: l'Ecole Polytechnique. In *The Counter-Revolution of Science: Studies on the Abuse of Reason*. Glencoe, Ill.: Free Press.

Hayes, J. E., and Michie, D. 1983. *Intelligent Systems: The Unprecedented Opportunity*. Chichester, England: Ellis Horwood.

Hayes-Roth, F., Waterman, D., and Lenat, D. B. eds. 1983. *Building Expert Systems*. Reading, Mass.: Addison-Wesley.

Hobbes, Thomas. 1651. *Leviathan, or the Matter, Form, and Power of a Commonwealth Ecclesiastical and Civil*. Reprinted in (Burtt 1939).

Hofstadter, Richard. 1955. *Social Darwinism in American Thought*. Boston: Beacon Press.

Horkheimer, M. 1947. *Eclipse of Reason*. Oxford: Oxford University Press. Reprinted New York: Seabury, 1974. Cited in (Weizenbaum 1976), p. 252.

Huxley, Thomas Henry. 1870. On Descartes' "Discourse Touching the Method of Using One's Reason Rightly and of Seeking Scientific Truth." Reprinted in (Huxley 1898), pp. 166-98.

———. 1874. On the Hypothesis That Animals Are Automata, and Its History. Reprinted in (Huxley 1898), pp. 199-250.

———. 1898. *Method and Results*. New York: Appleton.

Keir, M. 1917. Scientific Management and Socialism. *Scientific Monthly* 5: 359-67. Reprinted in (Pursell 1969), pp. 248-55.

Koyré, Alexandre. 1957. *From the Closed World to the Infinite Universe*. Baltimore: Johns Hopkins.

Lucas, J. R. 1961. Minds, machines, and Gödel. *Philosophy* 36: 112-27.

———. 1968. Human and Machine Logic: A Rejoinder. *British Journal for the Philosophy of Science* 19: 155-56.

Ludmerer, Kenneth M. 1972. *Genetics and American Society: A Historical Appraisal*. Baltimore: Johns Hopkins.

Masuda, Y. 1985. Computopia. In (Forester 1985), pp. 620-34.

McCorduck, Pamela. 1979. *Machines Who Think. A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. San Francisco: Freeman.

Minsky, Marvin, ed. 1968. *Semantic Information Processing*. Cambridge, Mass.: MIT Press.

Montaigne, Michel de. [1580] 1959. The Futility of Learning. In *In Defense of Raymond Sebond*, chap 3. Trans. A. H. Beattie. New York: F. Ungar.

Newton, Isaac. [1713] 1934. *Mathematical Principles of Natural Philosophy*, 2d ed. Trans. F. Cajori. Reprinted Berkeley: University of California Press.

Noble, David. 1977. *America by Design: Science, Technology, and the Rise of Corporate Capitalism*. Oxford: Oxford University Press.

Orwell, George, 1937. *The Road to Wigan Pier*. Reprinted New York: Berkley, 1961.
——. 1943. Looking Back on the Spanish War. Reprinted in *A Collection of Essays*. Garden City, N.Y.: Doubleday, 1957, pp. 193-214.
Papert, Seymour. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.
Pascal, Blaise. [ca. 1660] 1958. *Pensées*. Reprinted New York: F. Dutton. (Numbering follows the L. Brunschvicg edition).
Pursell, Carroll W., Jr., ed. 1969. *Readings in Technology and American Life*. New York: Oxford University Press.
Schanck, Roger C. 1984. *The Cognitive Computer: On Language, Learning, and Artificial Intelligence*. Reading, Mass.: Addison-Wesley.
Searle, John. 1980. Minds, Brains, and Programs; Author's Response. *Behavioral and Brain Sciences* 3: 417-24, 450-57.
Simon, Herbert A. 1981. The Psychology of Thinking: Embedding Artifice in Nature. In *The Sciences of the Artificial*, 2d ed. Cambridge, Mass.: MIT Press.
Stanton, William. 1960. *The Leopard's Spots*. Chicago: University of Chicago Press.
Taylor, F. W. 1911. *Scientific Management*. New York: Harper. Excerpts from pp. 5-7, 15-16, 38-47, and 116-19 in (Pursell 1969), pp. 240-48.
Turing, Alan. 1956. Can a Machine Think? Reprinted from *Mind* (1950) in *The World of Mathematics*, vol. 4, ed. J. R. Newman. New York: Simon and Schuster, pp. 2099-2123.
Turkle, Sherry. 1984. *The Second Self: Computers and the Human Spirit*. New York: Simon and Schuster.
Voltaire. 1764. Sect. *Dictionnaire philosophique*. Reprinted in *Portable Voltaire*. New York: Viking Press, 1949, pp. 195-99.
Weizenbaum, Joseph. 1966. ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM* 9: 36-45.
——. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco: Freeman.
Westfall, Richard S. 1971. *The Construction of Modern Science: Mechanisms and Mechanics*. New York: Wiley.