Man-Machine Systems Laboratory Report,
Department of Computer Science,
The University of Calgary,
CALGARY, Alberta T2N 1N4

The human as an information processor

by

David R. Hill

Research Report 85/214/27

The material In this paper was first presented as an invited contribution to the University of Calgary Humanities Institute Conference;

Images of Humanity
The implication of contemporary developments and discoveries in science for the uniqueness of the human being.

held at the University of Calgary, April 1984.

Summary

Human progress and uniqueness are both founded, in practical terms, on an ability to process information effectively. In less tangible, but equally visible terms humans have been distinguished by their humanity: such specifically human traits as altruism, self-sacrifice, love, and faith. It is no accident that benevolence is a synonym for humanity. In contemporary times, progress (in the guise of scientific discovery and development, and their engineering application) has, for some, eroded the notion of human uniqueness: (a) partly by 'explaining' the universe and apparently, thereby, denying the need for God and any special relationship He might have with humanity; (b) partly by explaining the the human as a purely physical mechanism lacking any intangible component or vital spark: and (c) partly by creating increasingly sophisticated replacements for supposedly unique human talents. This paper considers the relationship between humans and their artifacts, and attempts to justify the view of the human as an information processor. The paper goes on to examine the implications of that view for our ideas of human uniqueness in the contemporary world, especially a world in which the current epoch is described as the Information Revolution. An information revolution would seem to impinge directly on humanity's private preserve both by challenging human uniqueness, and by threatening to render humans redundant within the society they have created. The conclusions are that people continue to make exaggerated claims on behalf of science, including information technology, and that humans will continue to be unique as purposeful, spiritual beings. However, this is no reason for complacency. There is too much wrong with the human community. As computers do for brains what steam engines and the like did for muscles, we must bend our unique humanity to the benefit of humanity. If we fail in this we shall truly cease to be human.

Introduction

One of the distinguishing characteristics of being human has always seemed to be intelligence. The electronic digital computer, invented during the second world war and used for its ability to mechanise arithmetical calculations, was soon seen to be able to handle symbols and take decisions in a very general way. The computer was thus seen as a suitable basis for investigating the possibility of an artifical intelligence, a thinking automaton. By the second halt of the 1960's. Artificial Intelligence (Al) research flowered, as machines began to prove theorems in logic, learn about the world around them, understand and generate language (even in spoken form), play checkers and chess, and generally do things that were considered to require 'intelligence'. I was firmly committed to the field even then. One of the first things I did, on arrival as a new Assistant Professor at the equally new University of Calgary. was to inaugurate a graduate course in this new area. It was called non-numeric computing, to make it very clear that the object of the research was the use of computers in ways where the numerical ability *per se* was not the point at issue. Later the name was changed to Artificial Intelligence. when the term became respectable in Canada.

Being a good academic (despite my industrial research background) I felt the need to try to define the object of our study. I still feel the definition is fair. Intelligence is the ability to manipulate information effectively:

- Manipulate:
  - Search (heuristics)
  - Recognition (classification)
  - Selection (creation, evaluation)
  - Prediction (planning)
  - Acquisition (sensors)
  - Action (effectors)
  - Storage (remembering/forgetting, structuring)
- Information:
  - Accept Claude Shannon's definition, at least to start
- Effectively:
  - Achievement of the goals of the system: which implies goals
- Goals:
  - Desirable states/environmental-relations recognisable to the system
- System:
  - Collection of items whose intelligence is being judged

Intelligence is hierarchical, goal-oriented, and involves the manipulation of information, not passive transformation. It is essentially active.

Thus, if intelligence is an important distinguishing characteristic of being human, it is worth considering the human as an information processor. However, justice will not be done to this fascinating study in a brief review such as this.

Man, intelligence, and artifacts

One problem always faced in teaching computer science, especially in the areas of artificial intelligence, is that students and even colleagues can take a very narrow view of the human, based on a presumed understanding (at least at the conceptual level) of intelligence arising from the kind of definition noted above. Man's increasing understanding of the physical nature of the universe, and increasing dependence on knowledge as a basis for the 'good life', have tended to demystify creation (quite wrongly, I might add), and to suggest that the greatest good is intelligence, as understood in terms of an information processing model. Many go so far as to suggest that man is. after all, simply a biochemical machine run by a rather large and sophisticated computer. It is quite difficult to counter that view without mentioning faith, or God. or some other concept that sceptics will dismiss out of hand. Psychology has provided some of the characterisation of humans that has led to this unscientific and rather arrogant view and I am reminded of Cyril Burt's comment, quoted by Koestler in *The Ghost in the Machine* deploring an oversimplified view of the human engendered by psychological research based on the 'insights' of John Broadus Watson. Watson founded the behavioural school of psychology, which includes Skinner and Hull, and sees the whole complex of human behaviour as explainable in terms of directly observable elements of behaviour, formed in stimulus-reponse chains. The behaviourist view effectively purged psychology of any thoughts of consciousness, mind, imagination or purpose, and led to what Koestier calls "Ratomorphism"—the idea that man can be understood in terms of a rat in a Skinner box. One can understand the tenor of Burt's remarks:

> Nearly half a century has passed since Watson proclaimed his manifesto. Today, apart from a few minor reservations, the vast majority of psychologists, both in this country [UK] and in America still follow his lead. The result, as a cynical onlooker might be tempted to say, is that psychology, having first bargained away its soul and then gone out of its mind, seems now, as it faces an untimely end, to have lost all conciousness.

Only fairly recently have cracks begun to appear in the edifice that was put together under this regime.

Also in the 1960's. Joseph Welzenbaum created a computer program known as *Eliza*, after the character in Shaw's play *Pygmalion* who was moulded into an upper-class English gentlewoman by her mentor. Professor Higgins. Higgins wished to demonstrate to his friend that the upper classes were only distinguished from the lower classes by their accent and manners—superficial matters that could easily be manipulated and taught. which he proceeded to do. Higgins then fell in love with Eliza, as the protagonist did in the original classical Greek story summarised below. Weizenbaum's *Eliza* showed that a very simple approach to generating responses to natural language input could, in restricted situations, give the appearance of understanding. He regarded it as showing the way apparently intelligent behaviour could be simulated by a superficial mechanism, and was appalled when society fell in love with it. acclaiming it as a substitute for psychotherapy, and a model for an adequate substitute for natural language understanding. Psychotherapy came into the picture because, in order to help in hiding the machine's (program's) lack of understanding, the 'script' that Eliza followed was based on Rogerian psychotherapy techniques. A Rogerian psychoanalyst is able to ask questions as if he or she understands nothing, as a way of getting a patient to talk. Thus, part of a conversation between computer and patient might go as follows:

Patient: Men are all alike.

Computer: In what ways?
Patient: They are always bugging us about something or other.

Computer: Can you think of a specific example?
Patient: Well, my boyfriend made me come here.

Computer: Your boyfriend made you come here?
Patient: He says I'm depressed much of the time.

Computer: I am sorry to hear that you are depressed.
Patient: It's true, I'm unhappy.

Computer: Do you think coming here will help you not to be unhappy?
Patient: I need help, that much seems certain.

Computer: What would it mean to you if you got some help?
Patient: Perhaps I could learn to get along with my mother.

Computer: Tell me about your family.

Patient: ... ... father.

Computer: What else comes to mind when you think of your father?
Patient: Bullies.

Computer: Does that have anything to do with the fact that your boyfriend made you come here?

Eliza achieved surprisingly good results, with apparent flashes of insight, by a simple decomposition/re-assembly procedure, plus a few conversational tricks. A set of decomposition rules was selected from a 'script', the selection being based on the recognition of key-words. Thus the input: it seems to me that you hate me, might match the rule: 0 you 0 me. where 0 matches any number of words. The re-assembly rule might then be: Why do you think I 3 you, where '3' stands for the third component in the decomposition ('hate'). The conversational 'tricks' used knowledge about how people express themselves. For example. when a person says everybody does this or nobody does that they often have a particular person in mind. Other similar kinds of rules exist and this kind of knowledge can be built into the re-assembly rules associated with the decompositions resulting from recognition of the keywords involved. As there are several decomposition and re-assembly rules associated with each keyword, repetition is avoided, at least in the short term. If a keyword is not recognised, the computer has a number of stock phrases and ways of dealing with the situation. Two of these occur above. When the boyfriend is first mentioned, there is no match, and the phrase is simply echoed back. However, it is also placed in a 'memo' function (M). Later in the dialogue, when stuck for an answer again (in reply to *Bullies*), the computer uses the stored information, turned around by substitution of pronouns and such, with a re-assembly rule: *Does that have anything to do with the fact that M*. To those interacting with *Eliza*, the effect could be very convincing and there is the famous report of the secretary who asked Weizenbaum to leave the room when she was having her personal 'therapy'. However, if the conversation took a wrong turning, the effect could be quite the reverse:

Computer: I am sorry that you seem to be having problems.
Patient: Yes, things never seem to go properly for long.

Computer Can you be more specific about which things never seem to go properly for long?
Patient My grandfather clocks.

Computer Why do you think your grandfather clocks?

If the patient had said drinks instead of clocks then the reply would have made sense, in the context of psychoanalysing family problems, but the subject matter had misleadingly strayed outside the computer's knowledge. However, the computer was not programmed to know this.

Anyone who has ever travelled the Trans-Canada highway, along any forested section, will appreciate the aptness of a remark made by Hilary Putnam at a Conference in London (UK) in the mid-seventies. He said that some statements are nonsense in the way that the statement: "The number of trees in Canada is even", is nonsense. Subsequent to the reception of *Eliza*. Weizenbaum took issue with those who believe that knowledge based on this kind of "nonsense" can be programmed into computers to produce valid, or even useful output. He wrote a thought- provoking book (Weizenbaum 1975) pointing out that the inherent complexity underlying such deceptively simple statements, and the irrelevance of questions and answers framed in terms unsuited to the problems upon which such statements are presumed to shed light, result in computer programs that are incomprehensible, inflexible, innappropriate, and dangerous—although he does not use Putnam's example.

Society's reaction to the modest progress in applying computers alarms Weizenbaum, who sees the potential for dehumanisation, inflexibility, control, and oversimplification inherent in the unwise and over-hasty application of computers in areas we either do not understand well enough, or from which we should exclude computers for ethical reasons.

Much of the force of Weizenbaum's case derives from arguments about the level of understanding required to model situations or systems as a basis for solving problems; and from arguments about our ability (or more likely lack of ability) to implement such models as computer programs. Both sets of arguments centre on problems created by the complexity involved, as well as the character of the entities being modeled. These, in turn, affect the questions we ask, can ask, or should ask in order to formulate the model in the first place. From this ground. Weizenbaum argues that computers are being applied in harmful ways for a variety of derivative reasons. First, inflexible solutions to problems are created because complex programs—especially those written by a team—are themselves not understood well enough to permit changes to them, even to correct known errors. Secondly, solutions are based on incomplete models and data, due to our lack of understanding, and our lack of ability to formulate adequate questions to illuminate even those aspects we are aware of, let alone all the questions that we should ask, if we had God-like insight. There is also the question as to whether all relevant matters could be covered by such a factual approach. As riders to this, Weizenbaum points out that: (a) data may be ignored simply because 'It is not in the right form'; and (b) oversimplified solutions will be produced based only upon those aspects of the problems that we can formalise. A third harmful effect of computers, he argues, is that they act as a conservative force in society, partly by providing the means of sustaining outdated methods of running an increasingly complex society, and partly because, once programs are written, they are so resistant to change, for practical as well as economic reasons. Finally, he argues that computers have made society more vulnerable. With continued centralisation of control (itself outdated), errors and disturbances have far-flung and unpredictable consequences as they propagate through a homogeneous system, optimised for economy rather than stability. The scheduling of airline flights is an example of such a system in which unplanned hijacking incidents have propagated their dislocating effects on a world-wide scale, by domino action in a system with inadequate flexibility. It is also increasingly obvious that, as in all human activity, economic considerations tend to act in such a way as to simplify solutions and inhibit improvements that cannot be proved to bring directly measurable financial or political benefits. Such attitudes are much harder to attack when entombed in the amber of computer software.

Alongside this technical theme to Weizenbaum's book, there runs a strong philosophical argument against the dehumanisation of life and society. The most important point is this: by insisting that logical[1] solutions to problems are equivalent to rational[1] solutions to problems. one is defining out of existence the possibility of conflicting human values, and hence the human values themselves. Here can be seen the basis of conflict with many researchers in Artificial Intelligence, for the whole philosophical thrust of the book is against the view that the human being is just a computer, with mechanisms and rules that can be understood and transferred to a machine.

There are really two kinds of-questions raised by the book. On the technical side, there are questions about the best division of labour in a system involving both humans and computers: questions about the practicality, validity and utility of partial solutions to problems we do not fully understand; and questions about the state of our knowledge concerning how to implement certain kinds of solutions adequately. There is also the question as to whether some kinds of problems are amenable to programmed solution at all. These are valid research questions that cannot be ignored as we design increasingly complex systems. We should not get carried away by the modest success in improving knowledge access that has been achieved on the basis of rule-based "expert systems"[2].

The other kind of question begs the reader to step outside the conventional framework of disinterested science and ask questions about the value and ethics of what is being done with computers in terms of replacing people and running society. The underlying, but unstated message here seems to be that, if we are approaching God-like powers with our technology, we need God-like wisdom and restraint in the exercise of these powers. The implication is that the only viable basis for restraint and wisdom, on the scale required, is for each individual in the technological and scientific areas concerned to take some personal responsibility for the consequences of exercising his or her professional skills.

Pygmalion, from which *Eliza* got its name, was based on a much older classical story of a sculptor (Pygmalion. grandson of Agenor, king of Cyprus), who fell in love with an ivory statue he had carved. Aphrodite granted life to the image, and Pygmalio subsequently married his creation. Man has always had a fascination with the idea of creating artificial creatures endowed with human qualities, and has always tended to anthropomorphise his artifacts. People give their toys and tools names, and talk to them, chastise them, and beg them to perform. Albertus Magnus (in the middle ages), Bacon, Descartes and other philosophers built mechanisms in human form to open doors, or play music. Mythology is replete with legends of man creating monsters, automata, and homunculi. Frequently. the creations get out of hand and have to be stopped or destroyed, one way or another. An old example is the story of Rabbi Low's *Golem*, created to save the Jews in Prague from persecution in the 16th. century. The creation was modeled from clay and mute, since only God could give the gift of speech. It stopped the oppressors in their tracks, but finally ran wild, and had to be destroyed. The sorcerer's apprentice story, immortalised In our age by Walt Disney in *Fantasia*, is another example in which an automaton, created to help the apprentice with his work, got dangerously out of control until quelled by the sorcerer. The theme is a common one in science fiction, whether one takes Dr. Frankenstein and his monster, or HAL, the computer in *2001: a Space Odessey*. It Is a curiosity that if every letter in HAL is transposed forward by

1 Webster defines rational as having reason or understanding; being reasonable; whilst logical means formally true. Logic is, ultimately, tautologous, and denies conflict

2 See: "HILL DR (1994) Changes facing the university: our mission and how we may fulfill it"

http://www.cpsc.ucalgary.ca/~hill/papers/changes.pdf for a discussion of types of knowledge one letter, we get IBM—one of the major computer manufacturers of our present time. HAL was unable to reconcile the directive to keep the crew in the dark concerning the true nature of their mission with his normal logical functioning. The result was a homicidal computer that had to be 'killed', but this was done too late to save most of the crew. Since then HAL has been reborn, purged of guilty knowledge, and the ultimate fate of this electronic *Golem* is still in the air (or, rather, in near space).

Such stories seem to reveal man's fear of his own creations and power, and the dangers of meddling with matters too arcane for mere mortals. There is possibly some element of this, not to be lightly dismissed, in Weizenbaum's reaction to those who feel they are clever enough to turn computers to man's benefit—for this is the aim of the artificial intelligentsia. It is interesting that the classical test for success in creating artificial intelligence—the Turing test—is founded on the machine's ability to mimic human thought and behaviour, so that AI research continues the old tradition. If, when communicating remotely with some entity, the observer is unable to deduce whether he is dealing with a machine or a man, then the machine successfully passes the Turing test, and is deemed intelligent.

It is not clear to what extent such a test might depend on the human-ness of the human, rather than the cleverness of the machine. Clearly, the machine might appear far too clever at arithmetic, but could easily be programmed to slow down and make mistakes. The problem in designing machines for any such mimicry is that they can probably interpolate observed behaviour well, but previously observed behaviour will not necessarily extrapolate well to behaviour that the designers overlooked or did not know about. So far, acceptable performance has only been demonstrated within very restricted domains. *Eliza's* domain of acceptable mimicry was extremely limited; but, within the domain, it was quite convincing.

Chess playing was, for some time. considered a crucial test of intelligence. As a result. in 1968. during the first peak of activity in artificial intelligence research. Donald Michie (who had created a world centre of excellence in AI research at Edinburgh) made a bet with a then almost unknown British chess player (David Levy—who was also interested in AI) that within ten years a chess playing program would beat him. Subsequently others joined in the betting till the sum at stake exceeded David's annual income. David Levy went on to become an International Master. Nevertheless, in the final play-off at the 1978 *IFIP Congress* in Toronto, the first game went to the human, the second to the computer, and only the third game decided the result—in favour of the human. A return match, at the London (England) *Advances in Computer Chess* conference in 1984, still went to the human Levy, who won all the games that were completed. The machine, apparently, broke down more than once so that not all games were completed. In 1983, a program written by Ken Thompson of Bell Laboratories did well enough in competition with humans that it achieved the rank of US master.

AI research has made steady progress. The first real success was a program called Heuristic DENDRAL, which outperformed post-doctoral chemists in discovering the molecular structure of certain classes of organic chemical compounds. It was created by 'extracting' the relevant knowledge from those chemists in the form of basic algorithms. The DENDRAL algorithm itself could generate all possible molecular structures, given the atomic formula. The rest of the system is concerned with selection based on chemistry knowledge concerning mass-spectra, likely substructures, improbable structures and the like, so that the machine finally suggests only two or three alternatives structures as candidates for the true structure, usually with the actual structure at the top of the list. Heuristic DENDRAL was the first of a line of such systems (called Expert Systems since they embody knowledge culled from experts and then behave as substitute experts). A more recent example is a situation-action expert system called PROSPECTOR. This program uses geological and geophysical data to predict the occurrence of mineral deposits. and has already predicted one commercially important ore body that was missed by human prospectors. This kind of expert system stores its knowledge in the form of rules applicable to particular situations and is able to gather the information needed to apply the rules by dialogue with the user. Such programs are able to present a rationalisation of their requests for information. and their recommendations, in terms of the specific rules that are activated. In this sense they are sophisticated data-base access mechanisms, although they have powers of inference which allow them to deduce facts not stored explicitly in the data-base, and may easily detect possibilities that the human overlooks simply due to the tedium or complexity involved in search and evaluation. However, they can accumulate arbitrary amounts of practical, relevant knowledge, to the extent that it can be extracted from the heads of experts and cast in the form of situation-action rules. We are far from reaching the limits of their power.

Being human in the information revolution

The invention of printing was, as everyone has been assured many times, a very important step. It allowed, for one thing, mass distribution of knowledge. This was an important advance in communication. It is interesting to reflect on the knowledge that is in books. In a very real sense, a book contains some of the personality and expertise of the author. By writing a book, a more durable form of immortality is conferred than would be the case for oral tradition, handed down from generation to generation. Printing made this encapsulation more convenient, more effective, and more easily distributed.

The invention of the steam engine and the development of the modern continuously rotating lathe heralded the industrial revolution. Like the Rabbi's *Golem*, the lathe was developed and refined for purposes of war; it made for more accurate gun barrels. Lathes have been around in primitive form (in the East) since the earliest times, but the modern lathe was developed on the basis of machines making machines—probably the most important of Watt's inventions. A low quality lathe could make reasonable gun barrels, or a better lathe. By making a better lathe, better gun barrels were possible. The outcome, of course, was excellent lathes and excellent gun barrels. Strangely enough, major support for AI research in America comes from DARPA (formerly ARPA), an agency of the military establishment.

The industrial revolution had profound effects on society. Communication was further improved, machine power supplanted muscle power, and mass-production put more goods on the market and more money in the peoples pockets to buy those goods. Although the benefits took time to propagate, modern western prosperity finds its roots in the industrial revolution. However, there were also disadvantages. Society became more regimented and the working and living environments were fouled by the excrement of the new industries. Work became machine-like, repetitive and unsatisfying, giving rise to visions such as *Modern Times*, a movie in which Charlie Chaplin portrays a production line worker, who is caught up in the machinery in several different ways. He is even treated to the indignity of being force-fed by a feeding machine that gets out of control. The same period produced Fritz Lang's *Metropolis*, in which a mechanised utopia was seen to be founded on the degradation and virtual enslavement of a great mass of workers toiling in subterranean regions, sacrificing themselves and their children to the gods of the machine. In the midst of the plot there is again a mechanical creation (of the mad scientist Rotwang), designed to subdue the workers and (ultimately) deprive them of even their miserable living, to further enhance the profits of the master. The artificial human, without heart or soul, but otherwise a carbon copy of the lovely heroine, turns on her creators, and is destroyed—being revealed for the artificial creature she is, in the process. Surprisingly, all ends happily, but the messages are clear: greed enslaves; man becomes subservient to his creations which, no matter how clever he is. are quite likely to turn on him; and individual heroism and love—humanity—provide solutions to some of the problems man creates.

Art holds a mirror to life, and such films only caricatured the lives, hopes and fears of society. In reality, man began to interpret himself in terms of his creations, seeing himself as mechanical, and creating models of himself in the mechanical domain. The Czech play *R.U.R.* (*Rossums Universal Robots*) by Karel Kapek introduced the word 'robot' into the English language in 1923, four years before Fritz Lang completed *Metropolis*. It is was first used to describe mechanical devices so ingenious as to be almost human; or workers whom mechanical or repetitive work was making almost into machines. Robots, real or fictititous proliferated, but early real versions tended to be driven by steam, and capable of only mechanical effects; or (like von Kempelen's eighteenth century chess playing automaton) based on fraud, with a concealed human to provide intelligence. Von Kempelen also produced one of the earliest machines for producing synthetic speech. In the 1930's Westinghouse produced more than one robot, the best known being *Elektro*, who could walk, talk, smoke a cigarette, count on his fingers, and distinguish red from green. He was demonstrated at the 1939 New York World Fair, together with his dog *Sparko*, who could bark, beg, sit down and walk.

When I was a child, the automatic telephone exchange had become the latest technological innovation. Books purporting to describe the human nervous system to children likened it to such an exchange, with trunks and switches. It was convincing at the time. except that, like the chess player, it was a fraud. In the illustrations, little men sat at switchboards and consoles to provide the intelligence, including perception—just like von Kempelen's chess "automation". Even at that age, I could see the problems *that* caused. It was my first introduction to recursive problem solving and perhaps twisted my mind for ever, but it marked the beginning of my interest in information processing, as well as the mechanisms and nature of the human being. Such a false picture also highlights the one big weakness of all the artificial mechanisms up to that time. They lacked all but the crudest means of information processing. This meant that either they depended on human intervention, in some form, or they were not really human-like and could perform only the simplest tasks.

That has now changed. The industrial revolution is behind us, and the information revolution (so-called) has begun. There are interesting parallels with the industrial revolution, but, instead of enhancing or supplanting muscle power, the new machines enhance or replace brain power. Again man sees himself in his machines, and his machines in himself. The jargon of the computer world has become the *patois* of developed nations. Instead of books and mechanised printing, which preserve a factual, static image of their authors, we have dynamic information processes capable of interacting with their 'readers', and continuing to learn, whilst embodying something of the personality and character of their authors. These expert systems are superbooks, but will they fossilise our expertise by doing away with the need continually to regenerate and train human experts? The theme has been tackled in the modern mythology of *Star Trek*, and the answer is: possibly "yes"!

We have just reached a critical stage in the development of the revolution. We are on the verge of the equivalent innovation made by Watt, as a key factor in the industrial revolution—the machine to make machines, but these are information machines. Already our modern automata, of fact and fiction, can be endowed with a passing degree of intelligence. One can buy a box, little bigger than a paperback novel, marked with a checkerboard pattern, that will beat amateur players at chess. It costs less than $300. In the wake of David Levy's narrow victories, and the growing international stature of chess-playing computer programs, it now seems certain that we shall shortly see a machine take over as world chess champion.

If the industrial revolution had profound effects on society, the effects of the information revolution may well be greater. socially and technically. It was information processing that put humans in space and on the moon by providing the necessary world-wide control and tracking network. It was also a bug in a computer program that nearly lost the first lunar mission during the landing sequence, and it was the human who stepped in, took control manually, and saved the shore party. Similar kinds of bugs show up all the time, not only during computer program development, but also throughout the working life of the program. Another documented case concerns a bug in the autopilot program for F-16 fighters. Failure to take account of a change of sign in the variables, when crossing the equator, caused the automatic pilot to turn the aeroplane upside down at this point. Fortunately the bug was discovered during simulated flights.

The industrial revolution provided the basis for cheap mass war. Our weapons have grown ever more terrible, but the development of the atomic bomb depended heavily on the power of calculation brought to bear by the very first computers. Indeed, the development of the computer was stimulated, in the first place, by a need to calculate the ballistics of artillery shells in order to destroy the enemy more effectively. It is frightening to think that a bug in a computer program could start the third world war. Two documented cases representing near-misses exist. The first occurred because those programming the DEW line computers to detect Russian missiles rising over the arctic failed to allow for the presence of the moon. When the moon unexpectedly rose in a critical direction, the whole defense system nearly triggered. The second occasion occurred during war games. A tape was run through the computer to simulate an attack. Unfortunately, the computer reacted as if it was a real attack, and real missile systems began to respond. In the second case, human error was also involved, as it was at the Three Mile Island nuclear power station, where the human operators interfered in the automatic system that happened to be shutting down a faulty system correctly. The impossibility of building completely reliable systems. and avoiding human error, is the foundation of the case against the US Strategic Defense Initiative (SDI or "*Star Wars*") that is made by a large body of experts with relevant knowledge. It is sobering, in view of past events, that those with less knowledge dismiss such arguments out of hand. We truly begin to face a new *Golem*.

This kind of bad press, and ill-informed political control, adds to the fear of his creations that man has instinctively felt since the earliest of times. The combination of power, impersonality, and impenetrability that is engendered by computers touches a deeply buried chord in the human psyche. But now the homunculi and automata can be seen to have the power of thought. What then now distinguishes them from humans. With a few more years of explosive exponential development, what will be left for humans to claim as their own, to work at and excel in. Worse still is the thought that Weizenbaum's vision may prove true. Not only will the machine replace the human, but it will control the human in ways that are the essence of inflexible bureaucracy and unfeeling mechanism rolled into one. A science fiction story that catches a little of the flavour of this particular nightmare describes the trials of a library user who lost a copy of R.L. Stevenson's *Kidnapped*, and overlooked the fine. The system inexorably moved into gear, but without real understanding, and ultimately sentenced the hapless reader to death for involvement in the crime of kidnapping in which the victim disappeared (presumed dead).

It seems quite likely that one of the most important results of the growth in the home computer market will be the demystification of computers. A healthy first-hand appreciation for the strengths and limitations of the new machines will reduce fear and increase man's ability to control and use the new technology for good. It will provide the understanding needed to criticise and improve, and to resist the abdication of human responsibility. Far too often have we heard the excuse "I'm sorry, the computer made a mistake," knowing full well that it is the programmer or data entry clerk or designer who made the mistake. As Weizenbaum has pointed out, such mistakes may even, for some purposes, be unavoidable, and simply indicate that there are tasks that computers cannot or should not do. Weinberg apparently anticipated that one in 1967:

> Machines and human beings. In a way those who argue for the existence of tasks performable by people and not performable by computers are forced into a position of never-ending retreat if they can specify just what their task involves, then they admit the possibility of programming it on some machine. If they specify the task only according to procedures for recognizing it, they stand in danger of being fooled by a clever simulation. Worst of all, however, even if they construct proofs that a certain class of machines cannot perform certain tasks, they are vulnerable to the possibility of essentially new classes of machines being described or built Furthermore, they bear the burden of proof that people can indeed perform the task in question—where again they stand in danger of being fooled by a clever simulabon.

However, Weizenbaum's view is an effective counter to this view and, in any case, it is hard to imagine the difference between an effective human simulation of an ability, and a real human ability. It really boils down to devising sufficiently demanding tests of ability—precisely the problem in checking for bugs in computer programs. There are others on Weizenbaum's side. For example. Victor Rosenburg in *The Scientific Premises of Information Science* in 1974 said:

> There is no doubt that large computer systems can effectively handle information as data. But the development of even the most sophisticated information retrieval systems has not enhanced our fundamental understanding of the nature of information. In fact I would argue that the development of automated systems has inhibited this fundamental understanding. The demands of the computer, and the computer industry, have reduced much of our fundamental scientific effort to very sophisticated paper pushing. If, as I believe, the nature of human information processing is fundamentally different from machine information processing, then the development of digital computer systems becomes an obstacle to the undertanding of information and its use.

Another convincing statement of the problem was made by Heinz Zemanek in the same year in *The Human Being and the Automaton*:

> The scientific problem with the human being is that we can estimate how long a complete and perfect description of his body would be, but we are far from writing such a description. And while earlier decades could flatter themselves that they were coming close to it, today we are no longer certain that our new insights do not increase the length of the text faster than we are able to write it down. And even if we were making good progress towards an excellent systems description of the human body, the abstract description, the set of equations and strange symbol chains which science can deliver will be in hopeless contrast to the live man with whom we shake hands. The human being is the result of an automatic production process called natural growth, and the resulting product has a systems character

> in many respects: it has lines for the transportation of air and food and blood, and it has a nervous network for the storage, transportation and processing of information. For a good systems description of a human being we could not ignore the history of the individual nor the history of his ancestors. The hypothesis that so-called simple physical laws govern every event in the body may be the correct theory, but it tells us nothing about the body as a system. And, in fact, not only in the present state of physical science, but also in the foreseeable future the systems description of man will have to be highly imprecise. Once I coined a very simple sentence to illustrate this. What could a lover gain from the switching diagram of his fiancee? It could be argued that this is an unfair remark. But any operation of the computer is in fact a man/machine system and it cannot be made clear enough that this means a cooperation between a structure where the switching diagram is everything, and a structure where the switching diagram is unavailable, and if it were it would mean nothing.

Zemanek's is an elegant statement of the basic point underlying Weizenbaum's argument. Certain kinds of knowledge are unknowable in formal terms and can only be acquired by being part of the human race, and being raised as a human being under particular circumstances. Even humans are limited by their circumstances. Weizenbaum illustrates this by pointing out that an American family court judge would not be able to do a good job as a family court judge in Puerto Rico. He simply would not have the deep understanding of the culture and customs that comes from being raised in the society. It is a problem faced by any immigrant and is recognised in the laws of many countries that allow people to return freely to the land of their birth, regardless of the course their lives take.

Of course, our modern mythology seems aware of this problem. In the recent movie epic, *Star Wars* (from which President Reagan's Strategic Defense Initiative got its nickname), the robots and 'drolds share work and leisure with their human masters. They are slaves, but well treated, and it is a short step from there to emancipation. Does this mean that the final objection to equality can eventually be overcome? It is noteworthy that, at the 1985 *International Joint Conference on Artificial Intelligence*, in Los Angeles, one session was concerned with issues of personhood and responsibility with respect to computers. Already, people have experimented with raising lower primates as human children are raised, as part of the investigation (stated or unstated) into the uniqueness of man and man's apparently unique ability with language. Suppose (and this is itself quite a supposition, well highlighted by Zemanek), suppose that we could create a *Golem* of the required complexity, and add the gift of speech. Furthermore, suppose that this creature of our technology did not spring from the earth, fully armed like the soldiers Cadmus grew from dragon's teeth, but grew and matured from dependence to independence, in the bosom of some human family. Apart from asking if this would make any kind of sense, ask if the machine would be the equal of a human, for this is the fundamental question, so memorably summarised by Douglas Adams in *The Hitch-Hiker's Guide to the Galaxy*, concerning life, the universe, and everything. If the answer is that it would make sense, then what are we doing here. What is the purpose of love and justice, of heroism and self-sacrifice? If the answer is negative, then what is missing. If we can describe it, can we add it? This is the ultimate question that our increasing cleverness forces on us. When all is said and done, the view of man as an information processor is valid, and advances our understanding of man as a mechanism, but it does not really advance our understanding of what is is to be human at all; for to be human is to experience life as a human, with its conflicting values, choices, uncertainties, pain and joy; to love, to grow, to think of others before self, and to face good and evil.

Conclusion

Thus the view of the human as an information processor is an evolved view of man the automaton. If, as seems to be the case, we understand ourselves well enough at that level to make passably efficient imitations, to land our aeroplanes, plough our fields, write our routine correspondence, diagnose our ailments, and find our resources, we have come a long way. There is still more to do at the level of man as an information processor, and it will be a long time before I am sufficiently convinced of the judgement and pattern recognition abilities of machines to submit to brain surgery by an automaton. However, I can believe that some day the required state of knowledge will be achieved. This still leaves untouched the age old question, usually cast in terms of free will versus determinism.

If humans are mechanisms, pure and simple, however subtle, then it is hard to see that human behaviour, human achievement, is any more interesting than the unwinding of a clock. Determinism seems to win the day. We are determined to be born, to live a certain life, and to die, perhaps to continue in some other form and place according only to our inescapable destiny, and not our merit or sins. If so, then why do we reward the great and condemn the criminals—they are but mechanisms, and we should study how to mend or emulate them, not punish and reward them. If the claim is made that this would not be practical, or that reward and punishment are the tools for control—does this not deny determinism? It surely denies justice.

If humans are not determined in their behaviour, but have real choice, then from whence comes the choice. If it is not purely random, then what determines choice? A mechanism, an algorithm? Surely that is determined. If God chooses, then that is God's choice, not man's, and man is again reduced to a mechanism.

I can see no easy resolution of the question, of course. But those who claim that man is simply a rather complicated machine somehow miss the point. Such a view explains nothing, gives no basis for action, and says nothing about the existence or non-existence of God. It only says something about our progress in building useful tools, and seeing some aspects of ourselves in them. If, on the other hand, as human experience repeatedly indicates, to be human is to be more than a mere mechanism, then we must accept that the distinction lies beyond logical argument or proof. It is, however, continually revealed in the way that humans respond with *humanity* to their fellow beings.

References

WEIZENBAUM. J. (1975) *Computer Power and Human Reason.* W.H. Freeman