ARTIFICIAL INTELLIGENCE AND INDUSTRIAL ROBOTS: AN AUTOMATIC END FOR UTOPIAN THOUGHT?

JAMES FLECK

Technology Policy Unit, University of Aston, Gosta Green, Birmingham, England

Introduction

Ideas of artificial men or thinking machines have pervaded legend and literature from the earliest times (1). It is perhaps only in the last twenty years or so, however, that technologies such as industrial robots and artificial intelligence have been developed which appear to have the potential to realize these ideas (2). The expression of such ideas and reactions to them have been diverse, embracing both the brightest utopian and darkest dystopian themes, and discussions are found in many different contexts, ranging from myth to critiques of current technology.

Artificial human-like constructs tend to stand as symbols for scientific and technological knowledge in general, and, indeed, are often seen as the ultimate products of such knowledge. The earlier myths and legends did not, of course, refer specifically to science and technology, but referred more generally to knowledge of an arcane and dangerous sort, often magical or mystical in nature. Consequently, accounts and discussions of robots tend to raise long-standing and very general issues, with perhaps an added degree of urgency and immediacy lent by the anticipated or feared potential of the currently developing technology.

While discussions of robots are only rarely made the centerpiece of utopian and dystopian accounts, they are certainly very common components, largely on account of their symbolic importance and, indeed, are becoming an almost obligatory element. This essay, in discussing robots, therefore, is focused not so much on utopias *per se*, as on a particular common scientific element of utopian thought, albeit an element of general import. Robots and thinking machines, because of the existence of a body of literature often explicitly

Everett Mendelsohn and Helga Nowotny (eds.), Nineteen Eighty-Four: Science between Utopia and Dystopia. Sociology of the Sciences, Volume VIII, 1984. 189–231.

utopian in nature, and the presence of a newly established area of practical scientific and technological activity, offer a convenient arena in which to examine the relation of utopian thought with the dynamics of scientific development. This is the major aim of this essay. In addition, more specific observations on the ideas and attitudes towards robots and thinking machines will be made.

To these ends the images in the wider literature, especially science fiction, are explored in the first section and two major themes explicated; namely, robots as dangerous knowledge, and robots as a projection of man.

The next section provides a brief outline of the actual state of affairs with respect to the practical developments in the areas of artificial intelligence and industrial robots. The situation is further explored by an examination of five typical positions adopted by protagonists in and around artificial intelligence research. The overall picture which thus emerges is one of the articulation of a new cognitive space, mapping the differentiation of scientific and technological artefacts and their associated social relations. Moreover, there are clear indications of the emergence of a new utopia. In the following section, the occurrence of this new utopia is put into a broader context, and finally conclusions are drawn about the relations between science and utopias in general, as elucidated by the case of industrial robots and artificial intelligence.

Images in the Wider Literature

In this section I want to map out the structure of ideas pertaining to robots and artificial human-like constructs which is found in the general literature, as an indication of the ideas and attitudes prevalent in the wider society. Accounts are found in a wide variety of contexts including myth and legend, the utopian literature, science fiction and fantasy, as well as popularisations and critical discussions of science (3).

Most accounts are very far removed from the areas of practical scientific and technological developments focused on in this paper. Consequently the influence of such developments on the wider literature has to date been rather limited, as will become abundantly clear. It will also become clear that the general ideas, especially the utopian/dystopian tension, tend to pervade and structure all accounts of robots, even those given by practitioners themselves. This provides a basis for arguing that a positive utopian view of robots,

closely associated as it is with a strong belief in the desirability of scientific progress, has been of great importance in the development of industrial robots and artificial intelligence, at the very least as a powerful motivating, mobilising, and legitimizing influence.

The literature, however, is enormous and too voluminous to survey in full here: Patricia Warrick for instance surveyed some 225 short stories and novels in her book, *The Cybernetic Imagination in Science Fiction* (4). As well as identifying common themes she attempted to classify these writings, using a computer analysis based on 33 coded characteristics of each story, and arrived at three basic categories: the isolated system model; the closed system model which includes dystopian accounts; and the open system model which covers more optimistic 'speculative transforming' accounts. However, the association of 'closed' with dystopian, and 'open' with utopian is not altogether convincing. Warrick has perhaps too readily identified with a positive progressive view of cybernetics – as evidenced by her use of computer analysis and systems thinking. Non-trivial, clear cut categories are difficult to identify in the literature, while on the other hand certain themes do appear to be compelling and relevant to prevailing attitudes and views on robots (5). I will focus therefore on some such themes as a means of discussing the literature. A great many are found in the literature on robots, but two in particular stand out: robots as dangerous knowledge, and robots as projections of man.

Robots are frequently symbols for arcane knowledge, often seen as dangerous in nature; for instance the *golem* (from the Hebrew for a 'formless mass') was shaped from clay and brought to life by sacred rituals and utterances, in order to protect the Jews in times of danger, and, unless great vigilance was exercised, it would run amok (6). More recently, of course, the mantle of dangerous knowledge has been taken up by science and technology, with robots and thinking machines often being seen as the ultimate products of such knowledge. Mary Shelley's novel is particularly interesting in this connection as it appeals explicitly to the shift from mysticism to science: Victor Frankenstein goes on to study modern physiology and chemistry only after a thorough saturation in the works of the alchemists. It never becomes clear just what was the basis for the principle of life that "infused the spark of being into the lifeless thing", the monster that Frankenstein had assembled (7). But this knowledge certainly proves dangerous for Frankenstein, as

the monster inexorably causes the deaths of members of his family and household.

Of central importance in the notion of dangerous knowledge is the question of moral responsibility. While many accounts make reference to man playing or denying God, and draw upon the Faustian image of selling one's soul for material power, in others the morality is shifted from a religious to a social or political domain. The issue of scientists' responsibility for the effects of their productions is complicated in the case of intelligent constructs, it is often suggested, by the question of responsibility towards the creation itself. Frankenstein is torn by this dilemma but chooses the former responsibility over the latter, and turns away from making a female companion for his monster lest they breed and overrun mankind, despite the monster's impassioned pleas to his creator (8).

This example introduces a related theme found in both fiction and nonfiction contexts: a translation of the creation myth from religious to scientific terms, in which man acts as an evolutionary agent who brings into existence a superior kind of being. Samuel Butler developed this with an explicit appeal to Darwinian ideas. In Erewhon, all machines, no matter how elementary, were banned, because of the danger that they might evolve far more rapidly than biological organisms until they surpassed and could therefore replace humankind:

> there is no security against the ultimate development of mechanical consciousness, in the fact of machines possessing little consciousness now (9).

Indeed, one favorite robot fiction plot parallels Ludwig Feuerbach's critique of religion in that a computer brain, the image of man writ large, actually becomes God, or at least something effectively omnipotent (10).

The notion of dangerous knowledge inherently involves a dialectic between the beneficial and the negative potential of that knowledge. On the one hand, artificial man-like constructs are seen to take over all the drudgery, leaving human beings free to realize themselves in specifically human and creative ways. On the other hand, it is feared that the artificial constructs will usurp the prerogatives of human beings in both thought and action, leading to the consequent decline or stagnation of humankind. This tension pervades virtually all literature and thought about robots. It is not the mere existence of positive and negative views towards robots that seems to be important but the

essential indeterminacy between the good and the bad that is crucial. While the utopian/dystopian implications are not always outlined in full, it is very hard indeed to find any reference to robots that does not, at least implicitly, draw on this backcloth.

Signs of the utopian/dystopian tension can be discerned even in *Frankenstein*, despite the overall pall of gloom. There are some gleams of good in the monster which does show itself capable of sensitivity and love (11). Perhaps if it had been nurtured, rather than deserted by its creator these gleams would have shone brighter. The archetypical expression of the dialectic, however, is to be found in Karel Čapek's *R.U.R.* (Rossum's Universal Robots), the play that popularised the word 'robot' (derived from the Czech word 'robota' meaning 'serf', with connotations of forced labor (12). *R.U.R.* is a satire of capitalism; the robots are manufactured from artificial tissue (and would be termed androids in today's science fiction) as a source of cheap labor. They are made without emotions or a childhood, which are superflous from the viewpoint of productive efficiency. Inevitably the contradictions of capitalism emerge: the displaced workmen riot against the robots; robots are used to quell the riots; and finally they are used as soldiers in wars between countries. Meanwhile Dr. Gall, head of the Physiological Department of R.U.R., under the influence of Helena who is concerned for the wellbeing of the robots and wants them to have souls, has been experimenting with giving robots extra attributes, 'the physiological correlates' of the soul. These improved versions become leaders and organize the rest of the robots to revolt against the parasitic human beings. However, the robots fail to secure the secret of their own manufacture and appear doomed to extinction, after having already wiped out all of humankind. But in the last scene a glimmer of hope appears as Alquist, spared because he, too, works with his hands like the robots, witnesses the dawning of love between two of the modified robots. He enjoins them to go out into the world as a new Adam and Eve, and the play ends with the Biblical quote . . . "And God created man in his own image . . .".

This is the classic statement of the utopian/dystopian dialectic, but there are many others in the literature. A particularly chilling depiction of a robotic utopia is given by Jack Williamson in his science fiction novel *The Humanoid Touch* (13). Here the robots, the 'humanoids', rapidly surpass man's own achievements. Under the impulsion of the 'Prime Directive': "to serve and obey and guard men from harm" which is built into their central plexus, the

humanoids zealously protect all human beings from any conceivable source of harm, including knowledge about dangerous technologies, and ensure their happiness by administering a drug, 'Euphoride'. This suffocating utopia of complete material security and absolute happiness is seen as a horrifying hell by some human beings, who flee and hide from the humanoids, choosing to retain the freedom and dignity that comes from achieving results by their own unaided efforts and making their own mistakes. However, in Williamson's much earlier book, *The Humanoids*, the same dystopian view is initially projected, but the negativity is not maintained throughout (14). Certain human beings are able to escape domination by the humanoids and, indeed, can command and control them. This is because they are of superior moral character, free from ambition, greed, or the will to use technology and knowledge for destructive purposes, and who therefore cannot come to harm from their use of such knowledge. Moreover, even those unfortunates lacking such moral strength can be educated by the computer complex controlling the humanoids, so that they too can eventually be set free. Williamson plays upon the utopian-dystopian dialectic, evoking first one interpretation then the other, so that the reader is never quite sure where he is. His major theme is that technological progress demands a corresponding development in human nature, to avoid technological destruction.

Perhaps the most sustained attempt at portraying a positive view of a robot utopia is by Isaac Asimov. He consciously reacted against what he perceived as a generally negative view of robots and set out to build a realistic but positive view of the use of robots (15). At the core of his attempt lies his 'Three Laws of Robotics' which were first presented in 1941:

(1) A robot may not injure a human being, or, through inaction, allow a human being to come to harm;

(2) A robot must obey the orders given it by human beings except where such orders would conflict with the First Law;

(3) A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

These laws are written into the 'positronic brain' of every robot and are designed to avoid negative consequences. Yet even with Asimov's positive approach, dystopian possibilities emerge and are explored by him; for example, the sterility of the world of Solaria, where a very small human population is supported in conditions of material luxury by a host of robots (16). The

process of robotic support has gone so far that the human beings have come to avoid all physical contact between themselves, preferring instead the service of robots. This alienation from physical contact, with all communication by 'trimensional viewing', has resulted in the loss of most of the reason for living, and societal stagnation appears inevitable.

The second major symbolic role played by robots in the literature is as a projection of man. Many accounts are not so much about robots as explorations of specific, usually inadequate models or interpretations of man. Sometimes satire of current social conditions is involved, with the interpretation of man (or woman) projected as a robot, drawn from the assumptions enshrined in the social order taken to a logical conclusion. We have already seen this in the case of *R.U.R.*, and it is also evident in Aldous Huxley's *Brave New World*, with its mass-produced delta, gamma, and epsilon subhumans (17). While Huxley's subhumans were robotic by virtue of being made in several models for different purposes by modern process production techniques, the characters in Y. Zamyatin's *We* became robotized on account of their subjection to an extreme extension of Tayloristic regimentation and despotic control. Most of their day and all of their lives is centrally ordered and directed, so that they eat in unison and partake of sexual activity in accordance with a strict timetable and voucher system. Nevertheless, they are still human, and have been robotized only figuratively, until the new 'fantasiectomy' operation reduces their humanity even further (18).

Satire is pushed over into speculative horror by robotization in a literal sense, in a book by Ira Levin, *The Stepford Wives*, made into a film of the same name (19). Here men, successful in their careers in high technology industry or business, can join an exclusive 'Men's Association' in the small town of Stepford. This entitles them to have a robotic simulacrum of their wives made, with all the 'imperfections' – small breasts, bad temper, women's lib. tendencies, etc. – ironed out. The resulting 'ideal' wives are the ultimate products of consumer society: they love housework, talk about nothing but television adverts, are totally submissive to their husband's desires, and do not interfere in any way with the men's careers:

> That's what they all were, all the Stepford wives: actresses in commercials, pleased with detergents and floor wax, with cleansers, shampoos, and deodorants. Pretty actresses, big in the bosom but small in the talent, playing suburban housewives unconvincingly, too nicey-nice to be real (20).

The satirical intent, however, always rests upon the ultimate inadequacy of the robots and the essential difference of robotic character from the full open-ended richness and potential of human nature. This difference, in most robot stories, draws upon a dichotomous epistemology implicit in the Western cultural tradition since Greek times (21), which sees men and machines as fundamentally contrasted, along with mind and matter, subjectivity and objectivity, intuition and calculation, creativity and mechanism, free will and determinism, emotion and logic, soul and soul-less, and so on. One of the most convincing explorations of the differences between men and robots is in Philip K. Dick's novel, *Do Androids Dream of Electric Sheep?*, the subject of the film 'Blade Runner' (22). In this account, set in a desolate city-scape of the future, bounty hunters track down rogue androids. This is a difficult task since men and androids are very similar in all respects save for a crucial difference in their ability to show empathy. Ever more elaborate tests are devised to show up the difference, as the robot makers continually seek to perfect their products, and the situation can be complicated by the implantation of false memoires, causing androids to believe that they are in fact human. This crucial difference between human beings and androids becomes celebrated in a religious cult, Mercerism, which joins human beings together (as the androids can never be) in heightened empathetic awareness. In a quite literal sense then, in this story the robots are soul-less. In the film version the increasing similarity between robots and human beings is taken further, and the bounty hunter falls in love with one of the female androids he is tracking, and goes off with her/it.

In Asimov's stories the differences between robots and men are generally deliberately maintained by the robot manufacturers, with the robots being given metallic finishes and entirely logical and unemotional dispositions. Nevertheless, in some of his later stories, Asimov explores the possibilities of reducing or eliminating the differences. This idea is explicitly followed up in his 1976 novelette, *The Bicentennial Man*, where a robot, which has shown unusual signs of creativity (due apparently to some one-off chance variation in the manufacturing process) conceives a desire to become truly human (23). After undergoing many improvements to make it more and more like a human being, including a final one which takes away its machine-like immortality, it is declared by the world president to be a bicentennial man, on the two hundredth anniversary of his/its manufacture.

But underlying all of Asimov's stories is a more fundamental difference between robots and human beings. The Three Laws of Robotics, even though they codify the principles of good human morality, as Asimov likes to point out (24), are nevertheless imposed on the robots, thus restricting their free will. A human being who adheres to moral principles is good precisely because he has chosen to do so, and has rejected other possibilities of his own accord. Robots in contrast do not have any option as far as the Three Laws are considered and are thereby determined. Their behavior can never be moral because it is necessary. But the irrepressible Asimov, in one of his more recent short stories, 'That Thou art Mindful of Him,' even finds a way in which robots can escape from the compulsion of the Three Laws (25). In this story, robots are given a capacity for judgment, so that they may make the best choice under conditions in which, whatever they do, human beings will come to harm. Judgment enables them to decide whether to follow the orders of a child, an idiot, or a criminal rather than a roboticist; thus modifying the Second Law. This (inevitably!) leads to the robots so endowed finding that they themselves are the most fit to give orders:

> By the criteria of judgment built into ourselves, then, we find ourselves to be human beings within the meaning of the Three Laws, and human beings, moreover, to be given priority over those others (26).

Thus we see the difference between men and machines vanquished in yet another way, with rather sinister dystopian overtones.

Nearly every conceivable combination and permutation of themes centering around the implicit dichotomy between men and machines has been explored in the literature, even to the point of burlesque and parody. Pierre Boule (27) used the dichotomy in 'The Perfect Robot', where a brilliant professor designs, builds and perfects a breed of robots that can equal or better human beings in every conceivable way, yet they are not accepted as truly human-like, "lacking some undefinable characteristic" (28). Finally he solves the problem by introducing error and imperfection into the mechanisms of his robots. Then, and then only, are the robots recognized to exhibit fully human characteristics: they make mistakes and are imperfect – just like human beings.

Stanislaw Lem explores the dichotomy from the other side as it were, with his customary satirical humor and immense philosophical erudition, in *The*

Cyberiad – Fables for the Cybernetic Age (29). This series of stories features the constructors, Trurl and Klapaucius, who, the reader gradually realizes (for it is never made fully explicit) are robotic beings from a world set either in the future or in a wider cosmological context in which robotic intelligences play a full part. The logic in Lem's stories appears to be preposterous, but it is in fact a parody of the style of reasoning found in certain mathematical and philosophical contexts, particularly in automata theory (30). Lem's portrayal of a cybernetic world cannot be characterised as either utopian or dystopian, but rather as simply unexpected. In one story, for instance, Klapaucius finds a society of beings (the H.P.L.D.'s) at the Highest Possible Level of Development. Nothing much appears to be happening:

The plain shimmered beneath the square sun. Here and there, things stuck out of the sand, things like broken wheels, sticks, bits of paper and other rubbish, and the inhabitants lay any which way among them, one on his back, another on his stomach, and farther on was one with his legs up in the air. He wasn't a robot, but on the other hand neither was he a man, nor any sapient protenoid of the glutinous-albuminous variety. The head was round and plump, with red cheeks, but for eyes it had two penny whistles, and for ears it had thuribles, which gave off a thick cloud of incense. He was dressed in orchid pantaloons, a dark blue stripe down either side and appliqued with dirty scraps of closely written paper, and he wore high heels (31).

Throughout the *Cyberiad* there are constant digs at man's arrogant assumption of himself as the pinnacle of evolution, such as the above reference to "sapient protenoids of the glutinous-albuminous variety". This is taken to an extreme in 'Prince Ferrix and the Princess Crystal'. Princess Crystal has determined to marry none but a human being, a 'paleface'. Prince Ferrix, desiring her hand, is disguised as a paleface by the sage Polyphase who:

. . . took a blob of oily filth, dust, crud and varied grease obtained from the innards of the most decrepit mechanisms, and with this he befouled the princes vaulted chest, viley caked his gleaming face and irridescent brow, and worked till all the limbs no longer moved with a musical sound, but gurgled like a stagnant bog. And then the sage took chalk and ground it, mixed in powdered rubies and yellow oil, and made a paste; with this he coated Ferrix from head to toe, giving an abominable dampness to the eyes, making the torso cushiony, the cheeks blastular, adding various fringes and flaps of the chalk patty here and there, and finally he fastened to the top of the knightly head a clump of poisonous rust (32).

Even this disguise, however, could not match the 'genuine monstrosity' of a

real 'paleface'; but of course the Princess Crystal comes to her senses and marries Prince Ferrix after all.

Thus Lem devastatingly mocks the man is perfection school, who would hold that any attempt at constructing artificial intelligence or robots is misguided if not impossible. But at the same time he ridicules the opposite view that robots will be super intelligent and far superior to human beings. His cybernetic universe is 'peopled' by robots like the constructors Trurl and Klapaucius, who are all too fallible, petty and generally unreasonable, in short, human – despite their phenomenal intellectual powers.

The dichotomous epistemology, however, does not enter only into discussions about robots and artificial intelligence, but also emerges in other contexts and in particular in debates about the differences between male and female. Interestingly, it is feminine rather than masculine attributes that are emphasized in opposition to the machine, which becomes characterized as masculine, as is evident from Figure 1. Nevertheless, as Christine Woesler de Panafieu has pointed out, issues of gender are clearly important in the literature on robots and in early activity with automatons, expecially in France in the 18th and 19th centuries (33). Many recent stories feature female robots,

Man		Machine	
Intuition	(f)	Calculation	(m)
Irrational	(f)	Rational	(m)
Emotion	(f)	Logic	(m)
Mind	(m)	Body (f), matter (–)	
Feeling	(f)	Reason	(m)
Subjective	(f)	Objective	(m)
{ Freewill	(m)	Determined	(–)
{ Willfullness	(f)		
Soul	(m)	Soul-less	(f?)
{ Creative	(m)	Automatic	(–)
{ Procreative	(f)		

f = female
m = male

Fig. 1. Man/Machine and Male/Female dichotomies.

Of the attributes listed, only in the mind/body and soul/soul-less pairs is the masculine attribute emphasised in Man, and in those cases the attribution of a feminine characteristic to the machine is not strong. The categories of 'freewill' and 'creative' can be further differentiated into the stereotypical feminine 'willfullness' and 'procreative', but these are still distinct from the corresponding machine attributes.

and depend for their dramatic effect upon the contrast between the calculable and rational nature of robots, and the intuitive and irrational feminine stereotypes (34). A common theme is one in which a human male falls in love with a female robot (35).

The classic example (though not the first) in the science fiction literature, is the short story 'Helen O'Loy' by Lester del Rey (36). Two colleagues, Phil, a medical man, and Dave, a robot specialist, begin thinking about emotions and robots, and experiment with a domestic robot model, giving it mechanical glands – "complex little bundles of radio tubes and wires that heterodyned on the electrical thought impulses and distorted them as adrenalin distorts the reaction of human minds". Their experiments are only too successful, and the robot, Helen, falls completely and passionately in love with Dave. This naturally causes difficulties, but the two men do not have the heart to "yank her coils", especially when the robot protests to Phil:

> Don't, please! I can't think of myself that way; to me, I'm a woman. And you know how perfectly I'm made to imitate a real woman . . . in all ways. I couldn't give him sons, but in every other way . . . I'd try so hard, I know I'd make a good wife (37).

Dave eventually realizes he returns her love and accepts her as his wife. Phil never marries, however; for he, too, loved her.

No doubt, this theme is a projection of the male desire for an ideal female who is free from the 'unreasonableness', the deep and dark secrets of real human women, as Woesler de Panafieu suggests. At the same time, it is a reassertion of the belief in, and fascination with, scientific progress, personified (?) or symbolized by the robot, an extreme version of the male worship of machinery and technological artefacts such as the modern automobile (38).

Moreover, scholarship from a feminist perspective is demonstrating the relevance of gender ideology for the analysis of scientific development in general (39). The institutionalization of science involved the elevation of a particular form of rationality, which was sharply separated from ethics and opposed to feeling and intuition (40), a rationality clearly identifiable as masculine in nature. Consequently, the masculine characterization of 'machine', in the man/machine dichotomy is perhaps not surprising. Furthermore, Evelyn Keller has discussed how, as this institutionalization progressed in the 17th century, there was a marked polarization of the notions of

masculine and feminine, "in ways that would prove eminently suitable to the growing division of work and home required of early industrial capitalism" (41). In the light of this, the occurrence of stories about female robots, often specifically for domestic or leisure use, should not be unexpected. Nor is it perhaps surprising that today the point has been reached where, in both the man/machine and male/female cases, the validity of the stereotypes is crumbling and is threatened by replacement with newer conceptual categorizations. With the meshing, albeit imperfect, of the man/machine and male/female dichotomies, therefore, we see issues of more general import raised once again and the role that robots play in the literature as symbols for science and technology reconfirmed.

The distance of the accounts in the wider literature from present scientific and technological developments is in general all too obvious and underlines this symbolic role. There is a marked tendency to discuss the situation in terms of a full development of intelligence, nearly always embodied in an anthropomorphic form, if not as a thinking, feeling robot, then at least as a brain concerned with concepts and even desires in a wholistic human-oriented sense. The great gulf separating fictional robots from their real counterparts will become even more obvious after the brief outline of artificial intelligence and industrial robots in the next section. The model for the images in the literature is clearly man himself rather than any presently existing artefacts, even where authors have been informed or aware of these, as for instance with the writings of both Lem and Asimov. Warrick also makes this point after her wide ranging survey, and she suggests that "perhaps a new category – *futuristic* fiction – is needed, less supernatural than fantastic fiction, but less rigorously grounded in science than SF" (42). I would further observe that as far as robots are concerned, even accounts purportedly of serious intent and allegedly well based in science tend to indulge in rather wild assertions and speculations. I. J. Good's prediction of the first ultra intelligent machine by 1978 is a case in point (43). Another is Carl Sagan's description of the simulated robot in the language understanding program written by Terry Winograd, as an actual physical machine capable of real world perception, interpretation and manipulation, thus suggesting that a development still very far off from realisation was already in existence (44).

The Technologies and Their Development

After having discussed some of the general ideas about artificial human-like constructs, I will now briefly describe the practical state of the art in artificial intelligence and industrial robots, and give an outline of development to date. This task is made difficult because the images and ideas already in existence and in general circulation tend to pervade views about the current technologies, on the part of both supporters and critics.

Artificial intelligence is a computer-science related enterprise which emerged in the late 1950s, aimed at building computational models of aspects of so-called intelligent behaviour (45), usually in terms pitched at the conceptual level of thought itself, rather than at the physiological level of the processes underlying thought (46). The physiological level in fact provides the focus for attempts at modelling intelligence from the quite different perspective of cybernetics. Rather than being a substantive area oriented to the elaboration of a theoretical structure which comprises its subject matter, artificial intelligence has so far been based on a body of programming languages and practices, methods, techniques, and approaches. These have been passed on via the master/student apprenticeship process, and only recently have textbooks, formal courses and easily accessible exemplar programs become available. Theoretical work tends to be concerned on the one hand with the formal elaboration of this central core of methodology (47), and on the other hand with very speculative discussions about what constitute the major problems to be addressed. List processing languages, notably Lisp and more recently Prolog, are the basic tools for research, and are oriented to non-numerical uses, thus contrasting with conventional languages which are numerically oriented (48). This non-numerical emphasis, with a focus on logic and structure pitched at the conceptual level, is a crucial characteristic of artificial intelligence and plays an important part in the debates over whether work in the area is dehumanizing or not.

The central core of research tools is applied in a bewildering and increasing variety of application areas – vision, language understanding, medical diagnosis and metamathematics, to mention but a few. This efflorescence of research areas (schematically depicted in Figure 2) accompanies and reflects a move from the great optimism in the early days of artificial intelligence research during the 1950s and early 1960s, to a more differentiated state of

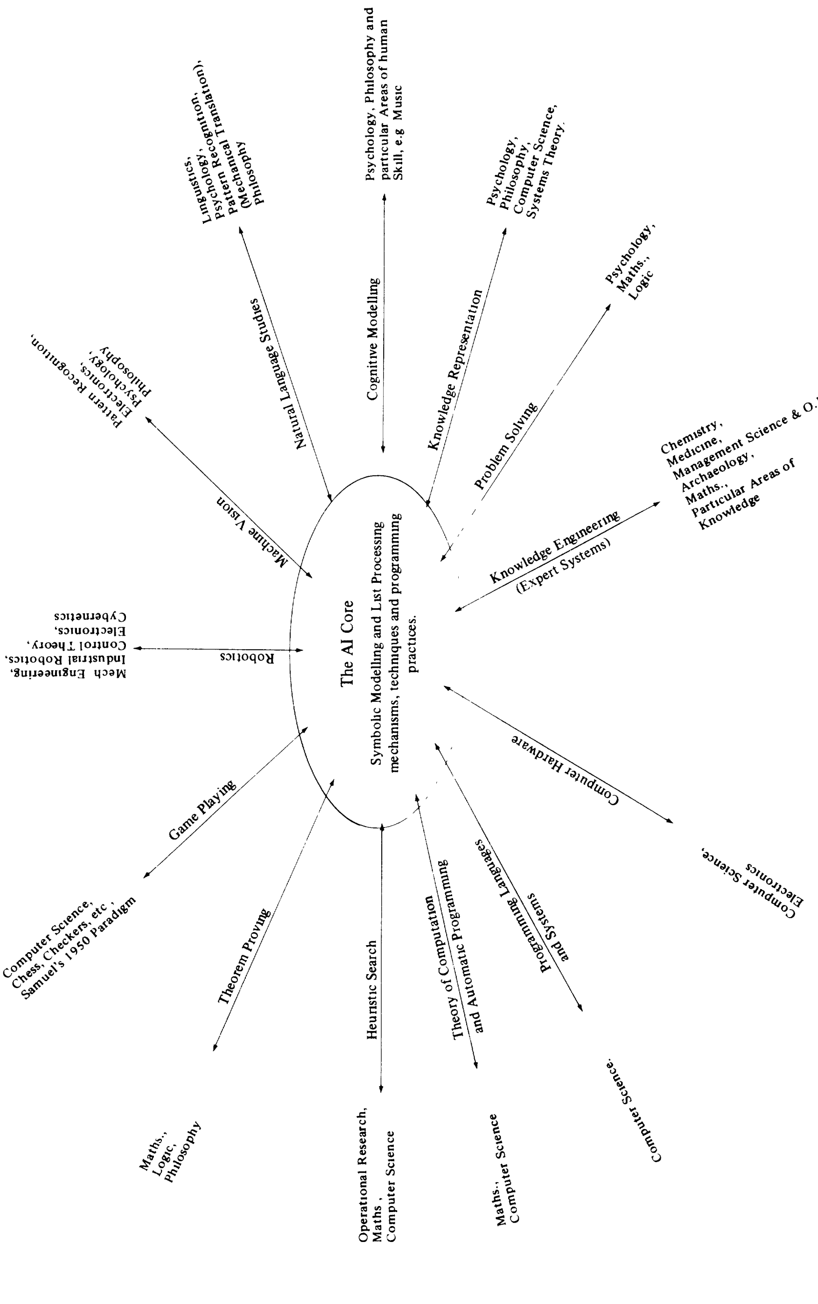

Fig. 2. Schematic representation of the structure of Artificial Intelligence.

affairs. Early optimism was evident, for example, in the attempts at Carnegie Mellon to build a 'General Problem Solver' (GPS), which could deal with any area of knowledge (49). After many years of effort during which the suite of programs comprising GPS became never more elaborate and articulated, reflecting the extent and complexity of the areas of knowledge concerned, the goal of a general intelligent inference mechanism was abandoned. In its stead a new goal – the representation of knowledge in restricted task domains – emerged. This approach explicitly recognized the specificity and contingency of knowledge with less emphasis on a simple correlation between intelligence and inference (50), altogether a very much more modest stance. A recent attempt to isolate a framework for representing and achieving 'intelligent' performance was given the acronym AGE – Attempt to Generalized – in recognition of this shift (51).

One of the most publicized application areas of artificial intelligence research is robotics, with intelligent robots, machine vision, and sophisticated assembly systems being investigated. Machine vision systems for robots have, in fact, already reached the stage of being marketed (52), but the artificial intelligence work is experimental, and not directly associated with the practical area of industrial robot technology which has emerged over the last 30 years. This technology emerged in the broader context of mechanisation and automation in production processes, as discussed by A. P. Usher, J. R. Bright and B. M. Bell (53), and summarized by R. Zermeno-Gonzalez as follows:

> Firstly, mechanisation is a process of refinement of machinery and of extensions to three fundamentally different tasks: transformation, transfer, and control. Secondly, refinement follows a trend towards the complete and continuous control of motion making the machine progressively independent of direct human intervention. This in turn demands the achievement of mechanical precision and adaptable control. Thirdly, mechanisation spreads to new tasks in a stochastic manner but steadily approaching the total integration of the factory (54).

The invention of the industrial robot is generally attributed to George Devol in 1954, but it was not until the early 1960s that the now standard industrial robot first became available on the market (55). The basic industrial robot configuration is that of an arm-like mechanical structure, free to move in unconstrained three-dimensional space (and therefore able to interface with a wide range of other machines) and with some control system that renders it autonomous. The arm is equipped with grippers or tools for carrying

out various tasks. There are also mobile robots which share the attributes of being free to move in space, and are autonomous, but these are not yet standard.

There is a wide range of robot-like devices (see Table I) as well as a wide range of robot applications: paint spraying, spot welding, arc welding, machine loading and unloading, palletizing, test and inspection, general handling, and assembly; but it should be noted that the total numbers actually in use are still relatively small: well under a thousand in the UK in early 1983 for instance, with very few indeed, several dozen at the most, in assembly, the most difficult area. Industrial robots vary in sophistication along several dimensions: the mode of construction, number of degrees of freedom; coordinate systems (according to the way in which movements along the 3 major positional axes are attained); control systems and the associated methods of programming; and the source of power. Different types of robots (there are

TABLE I
Robotic subgroups

Class of device	Autonomy	Structural versatility	Control versatility	
Servo-assisted hoists; tele-operator/telechiric systems for remote handling	None, require human operation	Varies	Supplied by human operator	
Pick and Place Devices (PPD'S)	Yes	Low	Generally low	
Playback robots	Yes	Varies low-high	Low-medium	INDUSTRIAL ROBOTS
Computer controlled robots	Yes	Varies medium-high	High	
Modular robots	Yes	Can be built to required level	Varies	
Mobile robots	Yes	Varies: free to move bodily	Varies	
Robotic systems	Yes	Varies	Varies	

several hundred models on the market) are best suited to different applications: cheap limited sequence, pneumatically operated devices are commonly used for plastic injection machine unloading, while revolute, hydraulically operated robots with continuous path control (i.e. one in which many points are sampled on a time basis to give a smooth play back of recorded movements) tend to find use in paint spraying. This latter type of robot, with a revolute configuration similar in appearance to an upside-down human arm, can make large fast movements in intricate patterns, but cannot achieve extremely high degrees of accuracy. Smaller revolute, electrically driven robots with more sophisticated computer control can achieve better accuracy and repeatability, and have been used for applications demanding closer tolerances such as arc-welding and simple assembly tasks.

As with artificial intelligence there were initially great hopes: The industrial robot was portrayed as a stand-alone general purpose machine (especially by J. F. Engelberger of Unimation, 'Mr robotics') (56), in distinction to other forms of automation, and there was believed to be great and immediate scope for widespread applications. In 1964, for instance, when GKN of the UK took up the licence for the American designed Unimate (the first commercially available industrial robot), it was thought that several thousand would be installed in a matter of months in GKN alone (57)! In the event, those hopes were soundly dashed – there were fewer than 200 units in use in the whole of the UK at the end of 1979, and in general diffusion everywhere was much slower than the manufacturers and promoters had expected, with only some 20–30 thousand robots in use worldwide by 1983 (58). Moreover, the purported general purpose character of industrial robots requires careful qualification. There has emerged a wide differentiation of robot types, as noted above, and practical experience has clearly demonstrated that certain robots are best suited to particular tasks. Thus any one robot embodies an element of flexibility within a narrow range of applications, rather than the full blown universality originally conceived. Furthermore, experience has also shown that the successful implementation of industrial robots is not a simple, immediate matter of plugging in a device bought 'off the shelf', but rather a lengthy process of robot applications engineering and organizational adjustment (59). Indeed, the growing body of robot applications knowledge, which is crystallizing from the collections of rules of thumb and the tacit knowledge gained from practical experience in the use of robots, is emerging as a newly

formalized and institutionalized element of production engineering, with robotics M.Sc. courses now on offer.

Clearly these industrial robots are far removed from 'thinking machines' or any notion of human-like robots, such as are found in the wider literature: their computational capacity does not extend beyond the calculation of relative positions and space-time trajectories, while the most sophisticated artificial intelligence component so far utilized in robotics, vision, is only rarely exploited. Likewise, with artificial intelligence research, very little progress has been made towards the full realization of an artificial thinker in a holistic sense. Progress – very dramatic in some cases – has only been made in extremely circumscribed terms. Very limited conversational exchanges can be sustained in very restricted domains of meaning, for example, and 'expert systems' provide a methodology for replicating the operational knowledge of human experts in specific areas, such as medical diagnosis or mineral prospecting. It appears highly likely that progress towards the unmanned factory or office of the future will come more from the continued rationalization of production processes and administrative procedures along Tayloristic lines, and their piecemeal automation, than from the development of man-like robots able to replace arbitrary workers (60).

But development in artificial intelligence and industrial robots cannot and should not be written off because little progress has been made towards a man-like robot (61). Despite the high flown rhetoric and continued speculation about the possibility of such developments in the future (62), this simplistic and utopian goal has been superseded as far as the practical scientific and technological activity is concerned. Instead, a differentiated set of more articulated and specific aims, with specialized knowledge and expertise developing around them, now structures and guides research and development.

This all adds up to the emergence of a new articulated area of knowledge covering the problem areas of artificial intelligence and industrial robots. New detailed problems and issues are being generated and created by the practical activity in these areas – problems and issues of which previously there was little or no conception or perception – and new structures of knowledge are being developed to explain and map them.

When the rather mundane practical developments are compared to the ideas and images in the wider literature there seems little point of contact. Nevertheless, high levels of motivation and enthusiasm on the part of practitioners

and widespread public interest have been a noticeable feature of development, in both artificial intelligence and robots, and without doubt have drawn upon the more positive utopian imagery for inspiration. The constant reference to Asimov's three laws and the now general use of the term 'robotics' coined by him, rather than the alternative 'robotry' offered in the *Oxford English Dictionary*, are perhaps indicative of this (63). Warrick also comments on the largely enthusiastic attitudes among computer scientists (64).

And yet, in fact, industrial robots so far have been slow to diffuse, their economic feasibility has been difficult to demonstrate, and robot manufacturers have found it hard to achieve profitability (65). Despite this, excitement still prevails and there is much activity, with well over two hundred manufacturers in what is a relatively small market (66). High levels of interest exist even at the national levels and industrial robots attract much more attention than numerically controlled machine tools, for example, which are arguably of far greater importance for the modernization of industry. This interest is based on the assumption that robots will be of great importance in the future. At the same time there is disquiet over the effects on employment, while newspapers and the management literature constantly refer to the fear of labor resistance, despite the fact that there is little or no evidence of any such resistance (67).

The evidence is clear, therefore, that the utopian/dystopian tension pervading the literature has also conditioned responses to the present generation of industrial robots. More importantly, it seems likely that utopian attitudes may well affect future developments through the self-fulfilling prophecy mechanism, by accelerating diffusion beyond that which would have been sustained on purely economic or technical grounds. Robots have become a symbol of national technological progress (68), a sort of international virility symbol, to such an extent that many companies have already introduced them without concern for the economics, to prove to themselves and others that they can handle new technology (69). Similarly many countries have grant schemes to promote the diffusion of robots (70).

Attitudes Toward Thinking Machines on the Part of Practitioners and Their Critics

Although a positive utopian attitude appears to prevail towards the current

simple industrial robots, there is also much concern, especially over the use of more sophisticated devices. In particular, there is a lively debate centering on artificial intelligence, which also covers robots of a more 'intelligent' variety. No one simple position is adequate to characterize the complicated situation obtaining here. I have, therefore, selected five typical positions which I think cover the range, and which also tend to be characteristic of the various groups in and around artificial intelligence research.

The five typical positions are:

(1) The simple utopian ideology of artificial intelligence.

(2) Artificial intelligence is in principle impossible and dehumanizing because it is reductionist – the simple dystopian view of artificial intelligence.

(3) Artificial intelligence is not impossible, but may be dehumanizing because it embodies an alien technological rationality.

(4) Artificial intelligence offers a way of humanizing technology, because it takes explicit account of human cognition.

(5) The approach embodied in artificial intelligence research has revolutionary potential, because it offers the possibility of transcending the dichotomous epistemology – a new utopia?

The Simple Utopian Ideology of Artificial Intelligence

Essentially, most artificial intelligence practitioners believe in the approach they are adopting. Some are certain that the approach is in principle completely all right, as was Alan Turing, the pioneer of the theory of computation (71), while others at least believe it is fruitful and worthwhile, and that the limits are a long way off. Occupying an extreme position, there are those, often hardware specialists or those most interested in the engineering applications of artificial intelligence and robotics, who hold the view that people are just sophisticated computers or 'meat machines', and who see nothing dehumanizing or necessarily reductionist about this: indeed, they often go on to discuss seriously whether computers will have the vote (72), when super-intelligent machines will emerge (73), and how they will then treat human bėings, their intellectual inferiors – one suggestion being that they will treat them as pets (74). In general, these people do not think that by replicating human behavior there will be a necessary belittling of what human beings can do; rather they tend to emphasize the positive possibilities, the potential for

removing drudgery and releasing people for creative activities or more leisure. This 'releasing people for more fruitful activities elsewhere' is a particularly common slogan in the promotional literature on industrial robots, whose main selling point is that they 'save labor' (75). The point that robots have, in fact, already led directly to redundancies and therefore unemployment in individual cases (76), is systematically glossed over, or ignored (77).

Clearly, then, this position constitutes a simple utopian ideology, a robotic version of the standard technological utopia, and substantial numbers of people subscribe to it. However, it is interesting to note that this utopian ideology tends to be most strongly held by newcomers to artificial intelligence or those on the margins. Furthermore, it was more characteristic of the area in general in the early days of research in the 1950s. Long standing and centrally involved practitioners tend nowadays to be more diffident and circumspect in their views, and while they may believe that ultimately robots will become social actors along with human beings, they are increasingly aware that moral and political issues are involved, and therefore can conceive that negative consequences may follow. For instance, Marvin Minsky, the 'meat machine' man, advised on the portrayal of HAL, the computer which took over a space mission and killed the human astronauts in the film *2001* (78).

Part of the reason for this shift in the state of affairs is that the standard artificial intelligence apprenticeship, the writing of a substantial computer program, generally involves the frustrating experience of wrestling for three years or longer to get the program to do something that people, even children, find easy. This greatly increases respect for human abilities and encourages an awareness of the complexity and subtlety of the issues involved. The process of goal differentiation and the efflorescence of research areas in robotics and artificial intelligence, already remarked, further increases this awareness, as many conflicting positive and negative consequences become obvious to the involved practitioners.

Another part of the reason, and one which follows the analysis of the development of new specialties given by W. O. Hagstrom (79), is that in the early days strong programmatic claims were made for research in the area, in order to win support and funding. With the successful establishment of the area, the need for such a utopia, as Hagstrom terms it, is diminished. In fact, nowadays one often finds concern over the dangers of overselling the

value and power of the products of artificial intelligence research (80). Newcomers and marginals, however, are not in a position to have developed an awareness of the complex and subtle issues involved, and therefore tend to go for the more simple and clear-cut utopian ideology.

But at the same time, there are many others, especially those in the humanities and those with phenomenological or holistic views of human nature, who view the prospects of the development of artificial intelligence and robots – for example the idea of human beings being treated as pets – as being profoundly dystopian. In a corresponding, but contrasting, position to the simple utopian extremists who believe a super intelligence will emerge, some believe that artificial intelligence is simply impossible in principle and that in any case the whole program is reductionist and dehumanizing.

The Impossibility of Artificial Intelligence and Its Reductionist Nature – the Simple Dystopian View of Artificial Intelligence

This is a very common position and many variants exist among critics of artificial intelligence, especially those with a philosophic bent and who are not involved directly with research in the area. It has been most extensively worked out by Hubert L. Dreyfus in his book *What Computers Can't Do: The Limits of Artificial Intelligence* (81). Dreyfus, a professor of philosophy, has been critical of work in artificial intelligence since the mid 1960s, and has taken part in many public debates.

Dreyfus explicitly focuses on the digital nature of artificial intelligence, the rule-governed character of programming, and what he sees as the undue formalization inherent in computational models. He argues from a phenomenological *Gestalt* perspective that these features are inadequate in principle for dealing with human thought and behavior. The brain does not operate on a digital basis, but an analogical one; hence the digital approach is wrong. Moreover, he believes behavior is not formalizable, because human beings do not, in fact, follow rule-governed information processes as used in artificial intelligence. Rather, there are uniquely human forms of information processing – 'zeroing in' in perception, or thinking with fuzzy ambiguous concepts in non-logical intuitive ways. Thought in human beings is mediated through, and is inseparable from, emotional aspects. Moreover, on ontological grounds artificial intelligence is wrong because the type of data computers

can deal with is restricted to the discrete, explicit, and determinate. Dreyfus is implacably opposed to artificial intelligence because he sees it as putting forward a view of man which misses what is essentially human; and also because in the long term, due to the malleability of human nature, people may in fact become like machines. It is perhaps surprising, but nevertheless consistent with his arguments that he does accept the in principle possibility of an analogically synthesized intelligence. This is a common feature among variants of this position, and appears to act as a counter to accusations of vitalism or religious fundamentalism – which, of course, are positions even more fundamentally opposed to the whole artificial intelligence endeavor.

Dreyfus has stirred up much protest in artificial intelligence circles, not the least because of his imputation of deliberate bad faith to leading artificial intelligence workers. He also arouses anger because he makes technical judgments and assessments that clash with the standards of practitioners in the area. They think he is simply incompetent and perverse in his interpretation of much work in the area, all the more so as he claims to have actually read and seriously considered that work, unlike many critics who base their judgments on a very detached understanding of what it is all about (82). An example, worth discussing because it is so often picked upon, concerns the digital basis of the computer and the implications of considering the analogical nature of the operation of the human brain. Allen Newell and Herbert Simon, leading artificial intelligence practitioners, commented on their approach:

> It can be seen that this approach makes no assumption that the 'hardware' of computers and brains are similar, beyond the assumption that both are general purpose symbol-manipulating devices, and that the computer can be programmed to execute elementary information processes functionally quite like those executed by the brain (83).

Dreyfus interprets this to mean that artificial intelligence people assume the brain operates by a biological equivalent of on/off switches (84). But Newell and Simon are explicitly rejecting this. They are saying there is only a *functional* similarity, which is shown in the end results concerned with modelling thinking.

Artificial intelligence workers also find it ironic that Dreyfus should attack artificial intelligence on *Gestalt* grounds, since they have explicitly appealed to such work as that of Edmund Husserl, in coming up with ideas for 'frames' and contexts of meanings (85). Moreover, as we have seen, artificial intel-

ligence work has moved increasingly towards emphasizing the importance and irreducibility of knowledge and skills, and is evolving methodologies which attempt to model such aspects in a structural and qualitative manner. While artificial intelligence practitioners are attacked for being reductionist, they see themselves as countering the greater reductionism of the dominant paradigm of behaviorism.

The view that artificial intelligence is in principle impossible and is reductionist thus comprises a simple dystopian view which contrasts digital and analogue, calculation and intuition, artificial intelligence and man, and draws on the dichotomous epistemology implicit in the wider ideological superstructure to distinguish man and machine. Although practical research in the area is explicitly considered, it is interpreted in a manner at odds with the practitioners' own conceptions and intentions, and strongly informed by the mind-machine images in general social currency and in the wider literature already reviewed. It is clear that it is a position that will be congenial to the wide ranging group holding epistemological commitments opposed to those held by artificial intelligence practitioners. Moreover, this group will tend to be non participating in, and indeed, highly critical of work in the area. It is also a long-standing antagonistic position. Michael Polanyi took up a similar position against Alan Turing's utopianism in meetings in Manchester in the 1940s (86).

However, the in principle impossible position is not the only distinctive one taken up by critics of artificial intelligence research. There is another well worked out stance that draws upon a similar tradition of thought as articulated by Herbert Marcuse, Jürgen Habermas and others, but applies it specifically to robots.

Artificial Intelligence Is Not Impossible, But May Be Dehumanizing Because It Embodies an Alien Technological Rationality

Here ideas of instrumental rationality as a source of domination are introduced, and the notion that artificial intelligence is impossible in principle is replaced by the idea that it constitutes an alien rationality which is thereby morally questionable. Moreover, the issue of a dystopian outcome derives from an overtly political process, having its roots in the domination of certain groups in society over others, a domination which, to quote Marcuse:

> ... generates a higher rationality – that of a society which sustains its hierarchic structure while exploiting ever more efficiently the natural and mental resources and distributing the benefits of this exploitation on an ever larger scale. The limits of this rationality, and its sinister force, appear in the progressive enslavement of man by a productive apparatus which perpetuates the struggle for existence and extends it to a total international struggle which ruins the lives of those who build and use this apparatus (87).

This sort of position tends to be taken by those who are close to the artificial intelligence activity, often being technically involved, so that they can appreciate the real power of some results, but who have strong doubts over their use and can see potential danger in their widespread application. They also tend to have a social and political awareness, in contrast to the two preceding positions which tend to be held by people who see the issues in strongly individualistic or psychologistic terms.

The position is well exemplified by Joseph Weizenbaum's book, *Computer Power and Human Reason* (88). Weizenbaum is a computer scientist who in fact has carried out some work in artificial intelligence, in language understanding, with a celebrated program ELIZA, the subject of many anecdotes. Weizenbaum was horrified to see how people read too much into this program and would even become emotionally involved with it, and this led him to discuss the dangers of work in the field. Weizenbaum sees the computer as a special case of instrumental rationality (as discussed by people such as Lewis Mumford, Jacques Ellul, and Theodore Roszak) which is often taken too far. Instrumental rationality turns everything into a technical issue, so the reality of human conflict and human values is denied. There is a tendency for scientific knowledge to delegitimize other ways of understanding, and the computer is a powerful public metaphor in which programmability becomes the only basis for explanation. The use of the computer fossilizes established institutions along conventional political lines and limits the kind of questions that can be asked, and the kind of data that is appropriate. It can have the effect of making people see men as machines, and the myth of technological, political, and social inevitability removes responsibility from those who believe it. Working with the computer can have dehumanizing effects on programmers: one finds a type – the compulsive programmer – who sees the whole world in terms of programming and pays little attention to other views of reality: programming becomes a self validating system of thought detached from human experience.

Weizenbaum does not say artificial intelligence is in principle impossible though he has doubts about its power. Rather he thinks that the impossibility of computers experiencing human socialization and development, where people come to know certain things as a consequence of being treated as human beings, means that computers, if they do develop intelligence, will develop an intrinsically alien intelligence. Moreover, he does accept as good some aspects of artificial intelligence. It shows up logical inadequacies in theories, can provide insight, and demonstrates how utterly primitive is our current knowledge of the human mind.

Weizenbaum appeals to the consciences of scientists, and asks them to choose not to do certain work, and to consider whether they ought to be doing what they are. In particular, he thinks two categories of work should not be developed: firstly, applications which arouse strong feelings, such as those involving interpersonal respect, understanding and love (for example, psychoanalysis (89)), and, secondly, those applications where there is a clear possibility of the occurrence of irreversible changes and undesirable side effects, especially when no pressing need is satisfied (for example, automatic speech recognition, which is expensive and therefore tends to be concentrated in the most powerful hands, with the attendant danger that it might be used for a repressive eavesdropping apparatus from which no-one could escape). This clearly constitutes an instance of the central inspection principle, a common element of dystopias, which originated as the panopticon of Jeremy Bentham, and is discussed by Michael Winter (90).

However, essentially the same structure of beliefs about the technical possibilities of artificial intelligence and robotics appears to be characteristic of other practitioners within the field, but interpreted in an overall utopian rather than dystopian form, in line with their interests as continuing and committed participants in the area. The position put forward by Donald Michie, professor of Machine Intelligence at Edinburgh University, in several of his articles and lectures is a convenient example here (91).

Artificial Intelligence Offers a Way of Humanizing Technology Because It Takes Explicit Account of Human Cognition

Michie argues that the instrumental rationality, the alien intelligence, is far more likely to come from mainstream computer technology and conventional

engineering than from artificial intelligence. These systems are usually based on formal mathematical models of bewildering complexity, so much so that human beings cannot master them. This is because they have certain data store characteristics (large numerical bases) and algorithmic characteristics (very complex ways of combining data and working out the answer) which lie outside the scope of human ability to comprehend in real time. The danger of such systems (for example, air traffic control) lies in the dependence upon them becoming so great, that even if there were suspicions that they had gone wrong, it would be impossible to intervene effectively, since to shut them down would precipitate even more accidents. He suggests that artificial intelligence, precisely because it does attempt to model and draw insights from the way human beings think, contrasts with the reliance on formal mathematics and numerical calculation of conventional systems, and can therefore provide interfaces between such systems and the human user, to enable effective control intervention to be made when necessary. One example of such a use is on a U.S. defense application, where a system had grown so large and complex, having been built up over a period of time by a number of people, that no one could understand it or operate it effectively. An artificial intelligence front end, designed to be an expert in the system, was used to mediate and enable human operators to use it effectively (92).

Recently, other such 'expert systems' have been developed (93). Indeed, the area has been heralded as one of the fastest growing new industries in the USA, and, thanks to the British Fifth Generation computer initiative, also looks set to grow rapidly in importance in the UK (94). Such systems are not restricted to interfacing with more opaque conventional systems, but, as already noted, have found use in previously non-automated areas such as medical diagnosis or mineral prospecting. There is even one system being developed to provide an expert guide through the British social security maze, so that claimants may easily obtain an understanding of their entitlement without having to run the gauntlet of bureaucratic officialdom. Michie himself is carrying out experiments to identify the 'human window'; that is, those particular data-store/algorithm combinations which human beings can comprehend, in order to make the design of such human-compatible systems easier. In short, we have here the view that, far from being an instrumental rationality which removes powers of decision from human control, artificial

intelligence can be used to humanize instrumental rationality, precisely because of the orientation to human ways of thinking.

Work in other areas of artificial intelligence is also seen as humanizing by the practitioners involved. One such area is the production of new computer programming languages and computing environments to enable easy access for everyone, and to remove the monopoly from computer experts. 'Programming for people' was a slogan used in this connection, and at Sussex University several years were spent in developing a computing environment congenial to social scientists and art students. All of this work draws explicitly upon the symbolic manipulation and structural modelling aspects of artificial intelligence. But this approach has met criticism from certain computer scientists who argue that it leads to sloppy programming and that programming is a discipline which is best left to the experts (for example, Edward Dijkstra, the structured programming proponent) (95). Here again we see positive and negative evaluations correlating perfectly with the competing interests of different groups.

A related position to the one just discussed is also in increasing evidence. It, too, involves the view that artificial intelligence is not necessarily dehumanizing, but goes further in providing a sophisticated articulation of the implications of work in the area. This position, just becoming crystallized, is largely held by people who do not have a primarily programming, engineering, or otherwise predominantly technical investment in artificial intelligence. Rather than being interested in the artificial intelligence programs and artefacts for their own sakes, these people are interested in exploring the implications of the associated models and views of mind for other areas of scholarship and scientific endeavor. There are now very many people from fields outside artificial intelligence beginning to take a positive and serious interest, and cognitive science is emerging as a major new area. Central to these developments is the belief that the computational metaphor has great potential to provide insight and a better understanding of how human beings think and behave, and a view that programming is one of the few approaches which can deal with complex processes and large inferential structures with a substantial degree of rigor.

These developments accompany the goal differentiation present in artificial intelligence and robotics and the development of a large range of artefacts, some of which are reaching the stage of practical industrial exploitation and

diffusion. All of this amounts to an explosion of interest in artificial intelligence which is reflected by the resurgence of articles and television programmes discussing the issues. In this context of rising interest, the approach embodied in artificial intelligence research is seen to have revolutionary potential, to such an extent that it can perhaps be seen as a new utopia.

Artificial Intelligence As a Revolutionary Approach – The New Utopia?

This position puts forward a sophisticated justification for artificial intelligence research in that it attempts to take into account humanist criticism. It has been furthest worked out by Maggie Boden in her book, *Artificial Intelligence and Natural Man* (96). Boden has a background in medical science, philosophy, and social psychology, and while she has done some programming work, prefers to use the medium of words. She recognizes the humanists' antipathy to artificial intelligence, appreciating that the idea of a mechanized mind threatens deeply held values and traditional beliefs. Boden suggests that the humanist dislike of applying psychological terms such as understanding, meaning, and so on to computer models stems at base from the fact that machines have no intrinsic purposes, but that machine purpose arises because of human design. She thinks, however, that the analogical use of psychological terms *is* justified in artificial intelligence. She suggests that it is a matter of decision whether their use is appropriate, not one of inspection of current usage, as some critics argue. However, she does recognise the ideological implications of language use (she uses the feminine pronoun for programmers and others throughout her book to good effect) and says there must be explicit acknowledgement of the analogical use, and priority accorded always to human ends.

Boden discusses the gap between mind and mechanism, where one can not translate theories using subjective concepts into non-psychological or mechanistic terms, but she suggests the universal employment of an internal representation of the program's world and goals makes it possible for a proper attribution of subjective terms in artificial intelligence. Artificial intelligence does not reduce subjective categories to physiological levels, nor see them as epiphenomenal, as do other approaches in cybernetics. Nor does it accept the Skinnerian behaviorist view which suggests subjective categories are at best

short hand labels, and at worst mystifying illusions. Instead, the internal model provides an analogue of subjectivity, and enables computer programs to generate structural complexities of performance comparable to the behavior of biologically evolved purposive organisms. Thus Boden believes that the image of machine provided by the artificial intelligence use of the computer can clarify the mind-body problem. She rejects the position of those, such as Dreyfus, who implicitly define intuition and other like terms in such a way that they are simply what the computer can't do, and believes, on the contrary, that recent work has provided interesting insight into human intelligence.

However, Boden does not insist that all aspects of human thought are in principle amenable to simulation. The epistemological issues involved are too obscure for any definitive claims to be made at present. Moreover, there is indeed a danger that the widespread use of computers could have bad effects and change our view of ourselves, removing the belief that we are truly purposive beings, and hence lead to the loss of an individual's sense of responsibility. But, on the other hand, artificial intelligence models are more markedly humane than the widely accepted and currently used behavioral views. The artificial intelligence view of intelligence as being expressed through computer procedures, with mistakes as program bugs which can be overcome, is, she argues, a more constructive and emancipatory view of intelligence than the view that it is the product of a number of mysterious, monolithic talents which one either has or forever lacks.

This sophisticated justification for artificial intelligence clearly has great motivational potential for practitioners in the area, and moreover, offers a legitimation for the far wider application of the artificial intelligence approach and ideas. It goes far beyond the range of concepts (which implicitly embody the dichotomous epistemology) generally available for discussing artificial intelligence and robots, by using terms pertaining to human characteristics in a modified materialist and realist manner. It thus raises the utopian hope for a resolution of the conflicts between objectivity and subjectivity, mind and matter, and man and machines (97). Furthermore, artificial intelligence has revolutionary potential not just because of its scope for practical application, but also because it has implications for the whole of the rest of science, by challenging the entrenched view of scientific objectivity (again based on the dichotomous epistemology). According to Boden it does this by:

Showing in a scientifically acceptable manner how it is possible for psychological beings to be grounded in a material world, and yet be properly distinguished from "mere matter". Far from showing that human beings are 'nothing but machines', it confirms our insistence that we are essentially subjective creations living through our own mental constructions of reality (among which science itself is one) (98).

There are undoubtedly many variants even of this sophisticated justification for artificial intelligence, and there are still extensive debates about the detailed arguments. Nevertheless, it does appear that a solid belief in the value and revolutionary potential of the approach is firmly established, and is spreading (99). While the extent of this belief is as yet still limited, and while there is clearly still an enormous amount to be done in the way of articulating and elaborating the full scope of the perceived potential, nevertheless there already exists the core of a broad program. Furthermore, there is already in existence a rapidly growing range of practical activities, an infrastructure, providing ample opportunity for the gaining of social experience to validate these new structures of thought. As a result it is perhaps justified to talk in terms of a new utopia.

Utopias, Industrial Robots, and Artificial Intelligence

Thus an examination of the structure of attitudes held by people closely located to artificial intelligence work reveals a complex pattern which is associated with the specialty positions of the proponents. Rather than one utopian/dystopian dichotomy sufficing to align those within the field against their critics outside, there is a relational structure. Various different utopian or dystopian interpretations tend to be emphasized by different groups of practitioners, and can be seen to relate to specific intellectual commitments and research interests. This complex social and cognitive structure, though it draws on the wider themes and ideologies found in the more general literature, is also a response to the specifics and contingencies of the actual scientific and technological development. Moreover, it is providing the context in which a new cognitive space is being opened up.

Accompanying the differentiation of research areas, the development of a wide range of artefacts, and the increasing interest and participation of many people from related disciplines, new patterns of thought are emerging to map out the more complicated relationship between human beings and

'intelligent' machines. From a relatively simple set of ideas and concepts reflecting the prevailing dichotomous epistemology, work in and around the area is beginning to generate an enriched vocabulary and an articulated interpretive structure for considering the relation of men to machines and the future possibilities for development. This extended cognitive space, as it percolates into the wider ideological structure presently dominated by the dichotomous epistemology of man versus machine, mind versus matter, intuition versus calculation, and subjective versus objective, will ensure that the advent of robots and artificial intelligence, no matter how limited or spectacular their capabilities, will be absorbed without either of the major simple utopian or dystopian outcomes being realized.

Among the articulations of this new cognitive space, there gleams the hint of a potent epistemology, the basis for a new utopia, waiting to take over from the existing crumbling dichotomies which have received so much sustenance from the three hundred year sway of the natural sciences. The issues raised by artificial intelligence and robotics are of central cultural importance, bearing as they do upon the conception of human nature and upon the relation of human beings to their material creations.

These issues have become crucial at the present time firstly because the monopoly of religion over the means of orientation with respect to the definition of human nature is being challenged (100), and secondly because we live in an increasingly industrialized and artefact defined and generated culture. Previous scientific developments have had wide ideological reverberations because the issues involved related to the question of man's place in the order of things, which meant at the time the relation of man to nature. Today the question of man's place in the order of things is becoming more and more the question of man's relation to his material products.

The central focus of many classical utopias was man's relation to nature, and natural laws, as revealed by science, became the ideal, with the emphasis heavily on the objective over the subjective, and matter over mind. This gave rise to the feared or hoped for assumption that human cognition could be reduced to natural laws; that the individual "is reduced to his calculable useful functions", while "utility becomes the supreme principle of the regimentation of things and human beings" (101); that humanity is "a law bound manipulable organism" (102).

These assumptions, which underlie the robotic dystopia in particular as

well as other dystopias in general, are, however, denied by the new utopia already outlined. The new approach is seen as offering a basis for understanding how human beings think or behave without reducing them to calculable functions. It is believed that the new approach can accommodate contingency, chance, and individual variability, without any intent to eliminate them. By challenging the man/machine and subjective/objective dichotomies, what is sought is not the extension of natural law to cover man, but rather the elimination of a purely instrumental conception of science and the reintroduction of mind, albeit on a material basis, into the operation of the material world. In this respect, the new utopia has something in common with the romantic utopian interpretations of high energy physics which seek to find room for subjectivity in terms of the contingency and uncertainty apparent at the elementary particle level (103). But with the new utopia of cognitive science it is not merely a matter of finding room for subjectivity within natural law, but the reassertion of law, as far as the behavior of human beings is concerned, as a medium for the interaction of moral agents, of conscious purposive entities.

Conclusion: Utopias and Science

It is clear, therefore, that utopias can influence and shape scientific development by providing motivation, legitimation, and even inspiration. This appears to be an inherently dynamic process. As scientific development proceeds, and social and cognitive differentation takes place, there may well be an increasing differentiation of positions, both positive and negative, which draw upon the societally available resources of utopian and dystopian images, but also reflect the research interests and commitments of those involved. The comments by Hagstrom on the motivational and legitimatory programs presented by proponents of emerging specialties, can be seen to fit in with these wider processes, as clearly these programs may be constructed using societally available resources of which literary utopias are a crystallization.

However, scientific and technological activity are constantly giving rise to results which go beyond the explanatory or accommodative capability of existing cognitive structures. Consequently new structures of understanding are always emerging to map out and explain the growing range of techno-

logical artefacts and scientific discoveries, and to provide a basis for describing and exploring the social relations of the new artefacts and the potential modes of production associated with them. In most cases it seems likely that the new cognitive structures engendered as a result of scientific activity will be effectively orthogonal to the utopian elements deployed, and will therefore have no direct nor easily identifiable effect on them. Of course, accumulations and combinations of events in the broader social structure will give rise to changes in the available resources, and to different utopian elements becoming predominant.

But occasionally it may happen that the scientific activity generates results with the potential for immediate and direct effect on the utopian elements, possibly by producing fundamental shifts in epistemological orientation, a process which is perhaps occurring in the case of the new utopia arising in association with artificial intelligence and industrial robotics. Such cases seem to evoke the dilemma discussed by Davis:

> . . . either scientific discovery remains free and utopia is exposed to the *fortuna* of its happenings and the radicalism of its consequences, or utopia is stabilised by controlling and ultimately stopping scientific discover (104).

However, the metabolism of utopian thought, if not the power of accommodation of particular utopias, appears quite equal to the challenge of regeneration posed by scientific activity. Indeed, is it not possible that Davis has got it wrong? Perhaps utopia and science, far from being the two horns of a dilemma, are instead complementary modalities of the same process. Perhaps Oscar Wilde was right when he wrote:

> A map of the world that does not include Utopia is not worth even glancing at, for it leaves out the one country at which Humanity is always landing. And when Humanity lands there, it looks out, and seeing a better country, sets sail. Progress is the realization of Utopias (105).

Notes and References

1. See H. M. Geduld and R. Gottesman (eds.), *Robots, Robots, Robots,* Boston: New York Graphic Society, 1978; J. Reichardt (ed.), *Robots: Fact, Fiction and Prediction*, London: Thames and Hudson, 1978; and J. Cohen, *Human Robots in Myth and Science*, London: Allen and Unwin, 1966.
2. Industrial robots and artificial intelligence have provided the most advanced ex-

amples of developments towards practical robotic constructs, although other areas of cybernetics and even genetic engineering are certainly of relevance as well. For a discussion of the potential of these other areas see: I. Aleksander, *The Human Machine: A View of Intelligent Machines*, St. Sapherin, Switz.: Georgi Publ., 1978; F. George, *Man the Machine*, London: Paladin, 1979; and D. Rorvick, *As Man Becomes Machine*, London: Abacus, 1975.

3. See the surveys referred to in Note 1, and also P. S. Warrick, *The Cybernetic Imagination in Science Fiction*, Cambridge, Mass.: MIT Press, 1980.
4. Warrick, *ibid.*
5. These themes are readily apparent in the surveys already noted as well as in many other reviews, such as 'the cybernetic society', Ch. 6 of J. Griffiths, *Three Tomorrows: American, British and Soviet Science Fiction*, London: Macmillan, 1980.
6. C. Bloch, 'The Making of the Golem', in Geduld and Gottesman, *op. cit.*, 1978 (Note 1), pp. 41–43.
7. M. W. Shelley, *Frankenstein, Or, The Modern Prometheus*, London: Dent, 1963 (originally 1818), p. 51.
8. *Ibid.*, Ch. 17.
9. S. Butler, *Erewhon*, London: Dent, 1963 (originally 1872), p. 141. For Butler's discussion of machines see Chapters XXIII, XXIV, and XXV of *Erewhon* and also an earlier piece, first published as a letter in 1863: S. Butler, 'Darwin Among the Machines', in Geduld and Gottesman, *op. cit.*, 1978 (Note 1), pp. 137–140. For 'non-fiction' accounts suggesting that man will bring into being a superior race of machine beings, see P. Davies, *Stardoom*, Glasgow: Fontana/Collins, 1979, pp. 164–168; R. Jastrow, *The Enchanted Loom: Mind in the Universe*, New York: Simon and Schuster, 1983.
10. See for example, I. Asimov, 'The Last Question', in his *Opus*, London: Granada, 1982, pp. 73–87. In this story computers become more and more advanced, as the entropy in the universe increases to a maximum. At this point the computer figures out how to restore everything and says "let there be light", and there was light.
11. Shelley, *op. cit.*, 1963 (Note 7), Chapter 11–15.
12. K. Čapek, *R. U. R.* (*Rossum's Universal Robots*), a play in three acts and an epilogue, in the Brothers Čapek, *R. U. R. and the Insect Play*, Oxford: Oxford University Press, 1961 (originally 1923).
13. J. Williamson, *The Humanoid Touch*, London: Sphere Books, 1982.
14. J. Williamson, *The Humanoids*, London: Sphere Books, 1977 (originally 1948).
15. See Asimov's comments in the introduction to his collection of short stories, *The Rest of the Robots*, London: Granada, 1968.
16. I. Asimov, *The Naked Sun*, London: Granada, 1960.
17. Aldous Huxley, *Brave New World*, Harmondsworth: Penguin, 1955, (originally 1932), p. 17.
18. Y. Zamyatin, *We*, Harmondsworth: Penguin, 1972 (originally 1920).
19. I. Levin, *The Stepford Wives*, in Omnibus edition, *Nightmares*, London: Michael Joseph, 1981 (originally 1972).
20. *Ibid.*, p. 199.
21. John Cohen discusses these roots – *op. cit.*, 1966 (Note 1); and also in *The Lineaments of Mind* (*in Historical Perspective*), Oxford: W. H. Freeman & Co., 1980. In

particular he notes that the Greek 'Automaton' is paradoxical in that it is used to connote both determinism and also acting of one's own will:

> "In the conception of a robot we seem to touch the limits of man regarded as fully determined by forces outside him, and also the limits of man as a creature of chance. We seem obliged to see him as autonomous or self-determining, as self-programming, as a computer scientist might say. This being so, man, we could provisionally conclude, is an *object* of study by others, but he is also a *subject* of study by himself of himself." *Ibid.*, pp. 27–28.

22. P. K. Dick, *Do Androids Dream of Electric Sheep?*, London: Granada, 1972.
23. I. Asimov, 'The Bicentennial Man' in his collection of stories *The Bicentennial Man*, London: Granada, 1978, pp. 164–207.
24. See for example, I. Asimov, 'Evidence' in *I, Robot*, London: Granada, 1968, p. 169:

> ". . . the three Rules of Robotics are the essential guiding principles of a good many of the world's ethical systems. Of course every human being is supposed to have the instinct of self preservation. That's Rule Three to a robot. Also every 'good' human being, with a social conscience and a sense of responsibility, is supposed to defer to proper authority; to listen to his doctor, to obey laws, to follow rules, to conform to a custom – even when they interfere with his comfort or his safety. That's Rule Two to a robot. Also, every 'good' human being is supposed to love others as himself, protect his fellow man, risk his like to save another. That's Rule One to a robot."

25. I. Asimov, 'That Thou Art Mindful of Him', in Asimov, *op. cit.*, 1978 (Note 23), pp. 79–105.
26. *Ibid.*, p. 104.
27. Pierre Boule is better known for his books *Monkey Planet* (1966), the basis for the film *The Planet of the Apes*, and *The Bridge on the River Kwai*, (1954), also made into a film.
28. P. Boule, 'The Perfect Robot' in Geduld and Gottesman, *op. cit.*, 1978 (Note 1), pp. 225–231.
29. S. Lem, *The Cyberiad: Fables for the Cybernetic Age*, London: Secker and Warburg, 1975. Lem's championing of robots versus men is even clearer in another collection by Lem: *Mortal Engines: Electronic Escapades in the lands of tomorrow*, New York: Avon Books, 1977. Consider also: "He began thinking about the innocence of machines, about how man had made them an accomplice of his mad adventures. About how the myth of the golem – the machine that rebelled against its creator – was a lie, a fiction invented by the guilty for the sake of self-exoneration." S. Lem, *Tales of Pirx the Pilot*, London: Secker and Warburg, 1980, p. 206.
30. For example, see C. Shannon and J. McCarthy (eds.), 'Automata Studies', *Annals of Mathematics Studies No. 34*, Princeton N.J.: Princeton University Press, 1956.
31. Lem, *op. cit.*, 1975, Note 29, pp. 253–254.
32. *Ibid.*, p. 286.
33. C. Woesler de Panafieu, 'Automates: A Masculine Utopia', draft paper presented to *Meeting on Science and Utopia*, ZIF, Bielefeld, Germany, Dec. 1982.

34. For example, I. Asimov, 'Feminine Intuition' in Asimov, *op. cit.*, 1978 (Note 23), pp. 15–40.
35. For example, L. Brackett, 'The Dancing Girl of Ganymede' (originally 1949), in A. H. Norton (ed.), *The Award Science Fiction Reader*, New York: Award Books, 1966, pp. 106–138; E. E. Kellett, 'The Lady Automaton', (originally 1901) in A. K. Russell (ed.), *Science Fiction by the Rivals of H. G. Wells*, Secaucus, N.J.: Castle Books, 1979, pp. 349–363.
36. L. del Ray, 'Helen O'Loy' (originally 1938) in Geduld and Gottesman, *op. cit.*, 1978 (Note 1), pp. 216–222.
37. *Ibid.*, p. 221.
38. This association is made explicit in several stories about robotic automobiles with feminine personalities. See R. Zelazny, 'Devil Car', in B. W. Aldiss and H. Harrison (eds.), *Decade the 1960s*, London: Pan Books, 1979, pp. 166–180, and I. Asimov, 'Sally', in *The Complete Robot*, London: Granada, 1982, pp. 7–24.
39. In particular, see B. Easlea, *Fathering the Unthinkable: Masculinity, Scientists and the Nuclear Arms Race*, London: Pluto Press, 1983. Easlea's use of Mary Shelley's *Frankenstein* as a basis for this analysis of the masculine nature of science, points up the symbolic importance of robots or man like constructs.
40. "The instrumentalisation of rationality, its separation from morality" was identified by M. Winter as one of the conditions leading to a negative utopia in 'The Explosion of the Circle', draft paper presented to *Meeting on Science and Utopia*, ZIF, Bielefeld, Germany Dec. 1982, p. 18.
41. E. F. Keller, 'Visions of Science Through a Feminist Lens', draft paper presented to *Meeting on Science and Utopia*, ZIF, Bielefeld, Germany, Dec. 1982, p. 20.
42. Warrick, *op. cit.*, 1980 (Note 3), p. 237.
43. Although admittedly perhaps not completely seriously: I. J. Good, 'The Social Implications of Artificial Intelligence', in *The Scientist Speculates: An Anthology of Partly-Baked Ideas*, London: Heinemann, 1962.
44. C. Sagan, 'In Praise of Robots', in Geduld and Gottesman, *op. cit.*, 1978 (Note 1), p. 167.
45. There are few adequate overviews of the area though textbooks are now becoming common, for example A. Bundy *et al.*, *Artificial Intelligence: An Introductory Course*, Edinburgh: Edinburgh University Press, 1972. For a discussion of the dynamics of development in the area in the U.K., see J. Fleck, 'Development and Establishment in Artificial Intelligence', in N. Elias *et al.* (eds.), *Scientific Establishments and Hierarchies, Sociology of the Sciences, Vol. VI*, 1982, pp. 169–217.
46. That is, in artificial intelligence programming the logical primitives of the program structures are themselves symbols standing directly for ideas or objects as conceived by people, rather than being quantitive potentials or neural patterns which embody the idea, or represent it indirectly.
47. For example, one of the theoretically most developed areas of artificial intelligence research is in theorem proving where there is an elegant body of mathematical theorems proving the adequacy of certain procedures.
48. The 'conventional' predominance of numerically oriented languages over non-numerical is associated with the historical use of the computer for number crunching work in high energy physics.

49. See G. W. Ernst and A. Newell, *GPS: A Case Study in Generality and Problem Solving*, New York: Academic Press, 1969.
50. This approach is embodied in the Expert Systems Methodology. See D. Michie (ed.), *Expert Systems in the Microelectronic Age*, Edinburgh: Edinburgh University Press, 1979, and D. Michie (ed.), *Introductory Readings in Expert Systems*, London: Gorden and Breach, 1982.
51. H. P. Nii and N. Aiello, "AGE (Attempt to GEneralise): A Knowledge Based Program for Building Knowledge Based Programs", *Proceedings of the 6th International Joint Conference on Artificial Intelligence 79*, Tokyo, 1979.
52. For example, the 'Univision' system used by Unimation based on a vision system developed from artificial intelligence research for the market by MIC, Machine Intelligence Corporation.
53. A. P. Usher, *A History of Mechanical Inventions*, Cambridge, Mass.: Harvard University Press, 1954; J. R. Bright, *Automation and Management*, Boston: Graduate School of Business Administration, Harvard University, 1958; and R. M. Bell, *Changing Technology and Manpower Requirements in the Engineering Industry*, Brighton: Sussex University Press, 1972.
54. R. Zermeno-Gonzalez, *The Development and Diffusion of Industrial Robots*, unpub. Ph.D. diss. University of Aston in Birmingham, 1980, Vol. I, p. 48.
55. *Ibid.* Also see J. F. Engelberger, *Robotics in Practice*, London: Kogan Page, 1980.
56. J. F. Engelberger, 'A Robot Factory Worker', *New Scientist* **29** (3 Feb. 1966), p. 270. There has been a continuing debate on the issue of more specialized systems with minimal robotic elements rather than stand-alone general purpose devices. See J. F. Engelberger, 'Stand-Alone VS. Distributed Robotics', in G. G. Dodd and L. Rossol (eds.), *Computer Vision and Sensor-Based Robots*, New York: Plenum Press, 1979, pp. 263–270.
57. Interview (by R. Zermeno-Gonzalez): R. F. Cakebread, Unimation Ltd., Telford, 29 Nov. 1978.
58. British Robot Association, 'Robot Facts, December 1982', BRA, Kempston Bedford, 1983.
59. See J. Fleck, 'The Introduction of Robots – Managerial and Organisational Problems', *Proceedings of the PEMEC 83 Conference*, Birmingham, June 1983, pp. 1A-4-1–1A-4-9.
60. Indeed, this has already been the case with those (nearly) unmanned factories now in existence, notably the Fujitsu-Fanuc factory in Japan and the 600 Group's SCAMP project in the UK. These still require human intervention for managing, monitoring, setting, and maintenance, despite the widely acclaimed 'unmanned' character.
61. It can be argued (and has been) that one of the most important contributions of artificial intelligence research so far has been the discovery that things which human beings find very easy to do, such as recognizing objects in complex messy situations and then picking them up and putting them together, are in fact extremely difficult to achieve using computers and robots. Ironically, it has proven far easier to successfully automate certain tasks which command great respect, such as some elements of mathematics or of chess – provided a sufficiently restricted context is employed.
62. For example the Horizon program, 'Better Mind the Computer', London: BBC TV, 21 March 1983.

63. Asimov himself suspects that he may well be remembered in the future for these contributions, rather than anything else – see 'The Time Travellers: Isaac Asimov is Interviewed by Christopher Evans', in G. Hay (ed.), *Pulsar 1*, Harmondsworth: Penguin Books, 1978, p. 81.
64. Warrick, *op. cit.*, 1980 (Note 3), p. XV.
65. For example, Unimation, the worlds leading robot manufacturer only made its first profit in 1975, and more recently on account of the recession and further investment exigencies went back into the red for the last quarter of 1982 and the first quarter of 1983. For an account of the factors facilitating and inhibiting the adoption of robots, see J. Fleck, 'The Adoption of Robots', *Proceedings of the 13th International Symposium on Industrial Robots, Vol. I*, Chicago, April 1983, pp. 1-41–1-51.
66. See the listing of manufacturers in the *BRA Members Handbook 1982/1983*, Bedford: British Robot Association, 1982.
67. Empirical studies find little evidence of labor resistance (see Fleck, *op. cit.*, 1983, Note 59) and this is borne out by the testimony of the robot manufacturers, notably J. F. Engelberger of Unimation who in his numerous addresses nearly always makes the point that problems of resistance derive from management rather than labor (he made this point in his keynote address to *Automan 83*, Birmingham, May 1983, for example). In fact, the enthusiastic reception given to the exciting new technology appears to prevail at the level of the workforce as well.
68. In fact robots have been used as indicators of national technological progress. See 'Study of the Adoption of Automation and Control Technology in the UK, Germany and Sweden', London: Systec Consultants Ltd., 1980.
69. This is surprisingly common. In 19% of cases of a study of adoption in the U.K., the decision to invest in robots was taken *before* a particular application had been found, and before therefore the economics could be appraised. J. Fleck, 'Robotics in Manufacturing Organisations', in G. Winch (ed.), *Case Studies of Technological Change*, London: Rossendale, 1983.
70. In particular the U.K. with various Department of Industry Schemes; Germany with schemes under both the 'Manufacturing Technologies' and 'Humanisation of Work' programs; and Japan with schemes directed or initiated by the Ministry of International Trade and Industry (MITI), such as the Japan Robot Leasing Company, (JAROL).
71. See A. M. Turing, 'Computing Machinery and Intelligence', *Mind* 59 (1950) 433–460.
72. N. S. Sutherland, 'Machines like Men', *Science Journal* 4 (*10*), (1968) 44–49.
73. I. J. Good, 'Speculations Concerning the First Ultra intelligent Machine', in F. L. Alt and M. Rubinoff (eds.), *Advances in Computers, Vol. 6*, New York: Academic Press, 1965, pp. 31–88.
74. E. Fredkin in BBC Horizon programme, *op. cit.*, 1983 (Note 62).
75. For example: "The employment of Unimates frees the human worker of drudgery, so that he can learn new skills and apply his latent talents to better ability in areas where a robot would be inadequate", in 'Industrial Robots: A Major Breakthrough in Automation', Telford, U.K.: Unimation Glossy Brochure, late 1970s, p. 2. It is worth noting here that cost justification of robotic installation is almost always worked out on the basis of pay back on labor savings alone.

76. See Fleck *op. cit.*, 1983 (Note 65).
77. For example:

> "But don't they take away jobs? Robots raise productivity and create wealth – which helps employment. Ignoring them has the reverse effect and, in the long term, can put jobs at risk. Remember that Japan has more robots than anyone, and less unemployment. There is little evidence so far that the introduction of robots has led to redundancies. The jobs that robots do are those that are boringly repetitive and are often in unpleasant – if not dangerous – conditions. They are jobs that are traditionally difficult to fill and in which turnover is high, and in some cases the first demand for the robot has actually come from the workers themselves. Employment is likely to be more secure in firms that employ robots, than in those that don't".

    in 'A Human Guide to Robots', London: Department of Industry and Central Office of Information, 1982, pp. 3–4.
78. Stanley Kubrik consulted many experts as well as Minsky: H. L. Dreyfus, *What Computers Can't Do: The Limits of Artificial Intelligence*, New York: Harper Colophon, 1979, p. 80.
79. W. O. Hagstrom, *The Scientific Community*, New York: Basic Books, 1965, pp. 211–220.
80. J. Mayhew gave a particularly passionate plea for due concern to the time constraints in what is a very difficult problem, in a presentation on Vision and Image-Processing at *Tutorial Meeting on Methods of Artificial Intelligence for Industrial Robotics*, IEE, London, Oct. 1982.
81. H. L. Dreyfus, *op. cit.*, 1979 (Note 78).
82. See the reviews of Dreyfus by artificial intelligence practitioners, for example, B. G. Buchanan, 'Review of Dreyfus', Stanford Artificial Intelligence Project, Memo AIM-181, STAN-CS-72-325, November 1972.
83. A. Newell and H. A. Simon, 'Computer Simulation of Human Thinking', the Rand Corporation, P-2276, April 20, 1961, p. 9. Quoted by Dreyfus *op. cit.*, 1979 (Note 78), p. 155.
84. Dreyfus, *ibid.*, pp. 155–156.
85. M. L. Minsky, 'A Framework for Representing Knowledge', in P. Winston (ed.), *The Psychology of Computer Vision*, New York: McGraw Hill, 1975, pp. 211–277.
86. For a later account of Polanyi's position see M. Polanyi, 'The Logic of Tacit Inference', *Philosophy* **41** (1966) 1–18. Andrew Hodges comments on the meetings in Manchester between Polanyi and Turing in his biography of Turing: *Alan Turing: The Enigma*, London: Hutchinson, 1983.
87. H. Marcuse, *One Dimensional Man*, London: Sphere Books (Abacus), 1972, p. 119.
88. J. Weizenbaum, *Computer Power and Human Reason*, San Francisco: Freeman, 1976.
89. Such models have been developed, notably by K. M. Colby, using a similar system to Weizenbaum's ELIZA: K. M. Colby, S. Weber, and F. D. Hilf, 'Artificial Paranoia', *Artificial Intelligence* **2** (1971) 1–26.
90. Winter, *op. cit.*, 1982 (Note 40).

91. For example D. Michie, 'Problems of the Human Window', talk given to 'AISB Summer School on Expert Systems', Edinburgh University, July 1979. Michie's views are distributed throughout his many articles and their concise expression is found in his oral lectures.
92. SACON, Structural Analysis CONsultant interfaced with a complex software package used in the design of aircraft wings. Reported by E. A. Feigenbaum at 'AISB Summer School in Expert Systems', Edingburgh University, July 1979.
93. See Michie, *op. cit.*, 1979 (Note 50), and M. A. Bramer, 'A Survey and Critical Review of Expert Systems Research', in R. D. Parslow (ed.), *Information Technology for the 80s*, London: Heyden and Sons, 1981, pp. 481–576.
94. The Alvey Report ('A Program for Advanced Information Technology', HMSO, 1982) identified Intelligent Knowledge-Based Systems (IKBSs) as one of four key areas, and proposed some £26m. over five years to develop it. Patrick Jenkin, the Secretary of State for Industry, in his statement to the House of Commons on the Government's intentions to substantially accept Alvey's recommendations commented: "... This is the first time in our history that we shall be embarking on a collaborative research project on anything like this scale. Industry, academic researchers and Government will be coming together to achieve major advances in technology which none could achieve on their own." See J. Alvey, 'UK Response to the "fifth generation" ', *Electronics and Power*, May 1983, 387–389.
95. E. W. Dijkstra, 'Programming: From Craft to Scientific Discipline', *Proceedings of the International Computing Symposium, 1977*, Liege, Belgium, April, 1977, pp. 23–30.
96. M. A. Boden, *Artificial Intelligence and Natural Man*, Hassocks: Harvester Press, 1978. See also M. A. Boden, *Minds and Mechanisms: Philosophical, Psychological and Computational Models*, Brighton: Harvester Press, 1981.
97. The resolution of these conflicts was also the hope of Norbert Wiener in *Cybernetics, or Control and Communication in the Animal and Machine*, Cambridge, Mass.: MIT Press, 1961 (originally 1948).
98. Boden, *op. cit.*, 1978 (Note 96), p. 473.
99. A burgeoning literature is witness to the excitement raised by the artificial intelligence approach. See D. C. Dennet, *Brainstorms: Philosophical Essays on Mind and Psychology*, Hassocks: Harvester Press, 1978; R. L. Gregory, *Mind in Science: A History of Explanations in Psychology and Physics*, London: Weidenfield and Nicholson, 1981; D. R. Hofstadter, *Godel, Escher, Bach: An Eternal Golden Braid: A Metaphorical Fugue on Minds and Machines in the Spirit of Lewis Carroll*, Hassocks: Harvester Press, 1979; M. de Mey, *The Cognitive Paradigm: Cognitive Science, a Newly Explored Approach to the Study of Cognition Applied in an Analysis of Science and Scientific Knowledge*, Dordrecht: Reidel, 1982; G. L. Simons, *Are Computers Alive*, Brighton: Harvester Press, 1983; D. Sleeman and J. S. Brown (eds.), *Intelligent Tutoring Systems*, New York: Academic Press, 1982; A. Sloman, *The Computer Revolution in Philosophy: Philosophy of Science and Models of Mind*, Hassocks: Harvester Press, 1978. Hofstadter has also introduced artificial intelligence ideas, including an outline of the LISP programming language, in his *Scientific American* column 'Metamagical Themas' (an anagram of 'Mathematical Games', the title of the column when produced by Martin Gardner), *Scientific American*, Vol. 244 (Jan. 1981) – Vol. 249 (July 1983).

100. See Fleck, *op. cit.*, 1982 (Note 45).
101. Winter, *op. cit.*, 1982 (Note 40), p. 10.
102. J. C. Davis, 'Science and Utopia: The History of a Dilemma', this volume, p. 21.
103. Cf. F. Capra, *The Tao of Physics*, London: Fontana, 1976.
104. Davis, *op. cit., this volume*, p. 21.
105. O. Wilde, 'The Soul of Man under Socialism', *The Complete Works of Oscar Wilde*, London: Collins, 1966 (originally 1895), p. 1089. This quote follows a passage describing the positive utopian view of machinery as slaves for man.