SOCIOLOGY Vol. 19 No. 4 November 1985
557 – 572

WHY NOT A SOCIOLOGY OF MACHINES? THE CASE OF SOCIOLOGY AND ARTIFICIAL INTELLIGENCE

STEVE WOOLGAR

Abstract In the light of the recent growth of artificial intelligence (AI), and of its implications for understanding human behaviour, this paper evaluates the prospects for an association between sociology and artificial intelligence. Current presumptions about the distinction between human behaviour and artificial intelligence are identified through a survey of discussions about AI and 'expert systems'. These discussions exhibit a restricted view of sociological competence, a marked rhetoric of progress and a wide variation in assessments of the state of the art. By drawing upon recent themes in the social study of science, these discussions are shown to depend on certain key dichotomies and on an interpretive flexibility associated with the notions of intelligence and expertise. The range of possible associations between sociology and AI reflects the extent to which we are willing to adopt these features of AI discourse. It is suggested that one of the more important options is to view the AI phenomenon as an occasion for reassessing the central axiom of sociology that there is something distinctively 'social' about human behaviour.

Introduction

The rationale for an examination of the relationship between sociology and artificial intelligence (AI) is threefold. Firstly, the recent growth of AI has been accompanied by renewed interest in the possible contribution of the social sciences to this area. In the U.S.A. and U.K. there have been calls for efforts to match the Japanese programme for the 'fifth generation' of artificially intelligent machines (for example, Feigenbaum and McCorduck, 1983). In the U.K. alone, a massive level of state intervention in information technology has spawned several projects attempting to apply social science perspectives to the development of information technology. In this context, it is both important and timely to clarify some basic assumptions about the relationship between social science and AI. Secondly, despite the evident rise in the AI phenomenon, it has been paid surprisingly little attention by sociology, compared with other social science disciplines. Thirdly, in the efforts of AI, several significant philosophical pigeons are coming home to roost. If, for example, AI turns out to be a feasible project, this would vindicate those philosophies which hold that human behaviour can be codified and reduced to formal, programmable and describable sequences. The perceived failure of AI, on the other hand, would amount to a victory in the eyes of humanist anti-reductionists. The positions adopted in the longstanding and fundamental debate about the nature of human action – whether or not humans are essentially rather complicated machines, whether social study should proceed on the basis that the human being is really just an animal and so on (for example, Pratt, 1978) – all stand to be revised or modified in the light of the outcome of the current massive research effort in AI. Clearly, it is important to develop an understanding of the social factors which will shape this outcome.

Since the literature on and about AI is vast, encompassing contributions from diverse perspectives within several disciplines, the discussion here is restricted to key features of current debates, with particular reference to 'expert systems'. This paper assesses the prospects for a *sociological* perspective on AI by drawing upon developments in the social study of science. The counter-intuitive sense associated with the notion of a 'sociology of machines' is reminiscent of the way in which a 'sociology of science' appeared problematic before Kuhn. How can a machine (science) be the object of fruitful sociological inquiry? An earlier response was to examine social relationships between scientists rather than the social character of scientific knowledge. However, post-Kuhnian sociology of science has made considerable progress in overcoming earlier preconceptions about its object of study and in modifying its own analytic perspective. Recent work has established that our understanding of science need not be so restricted; the nature and content of scientific knowledge is now recognised as a legitimate sociological object. By analogy, this paper examines the prospects for an approach to AI which construes machines as legitimate sociological objects.

The argument below is based upon a survey of discussions by a variety of sources about AI and about expert systems in particular. Three striking features of these discussions are evident. Firstly, in most of the literature, sociology is either excluded altogether, or granted an impoverished role by virtue of the use of a restricted notion of 'social'. Secondly, parts of the literature exhibit a marked rhetoric of progress in depictions of the likely development of AI. Thirdly, however, there is a wide variation in discussions and descriptions of the current state of the art. By drawing upon themes in the social study of science to account for these characteristics of the AI literature, it is possible to discern the basic concepts, distinctions and dichotomies which sustain current attitudes towards AI. This analysis then enables us to clarify the range of possible sociological perspectives on AI.

The restricted view of sociological competence

Although the associations between AI, on the one hand, and disciplines like psychology and philosophy, on the other, have been widely recognised and lengthily debated in terms of the implications of one for the other,[2] the important possibility of an association between AI and *sociology* has hardly been noticed.[3] On the few occasions it is alluded to, sociology is assigned the role of dealing with matters left over by other disciplines. For example, Boden's (1977: 446) notion of 'social' refers to the uses of artificial intelligence, the effects of futuristic proposals embodied in AI and the possible precautions to be taken by the AI community to minimise the dangers involved. Writers like Boden thus concede the relevance of the 'social' through avowals of concern for social responsibility. In this view, 'social' has to do with the *effect* of artificial intelligence, but not with its *genesis*.

Some sociologists have adopted a similarly restricted notion of 'social' in their treatment of AI (for example, Turkle, 1982; 1984). Although this can provide a suggestive characterisation of the environment in which AI takes place and of the way ideas about the computer enter into social life, it largely exemplifies the assumption that a 'sociological' approach to the phenomenon of AI entails the treatment of topics such as social attitudes to AI, public perceptions and acceptability of machine intelligence, and the likely effects of the implementation of AI in different institutional

environments (especially education). The contribution of the sociologist is seen to lie in discussions of the 'impact' of AI research, rather than in a detailed consideration of the process of the research activity itself. The danger of this restricted view of the scope of sociological investigation is that social science inputs into AI research are narrowly conceived in terms of investigations of cognitive psychology into human learning, memory, cognitive development and so on (Murray and Richardson, 1983). It is therefore important to understand the basis for this restricted view of sociological competence.

The rhetoric of 'social'

It is now generally accepted that post-Kuhnian attitudes to science embody a major revision in epistemological preconceptions about the nature of science (for example, Mulkay, 1979). While science was generally regarded as exotic and esoteric, it was neither necessary nor desirable for the sociologist to penetrate the content of science. In this 'received view' new scientific knowledge was assumed to be the rational extrapolation from the existing body of knowledge. Discussion of the content of science was largely limited to the analysis of the social origins of erroneous science; only when things went wrong was it the sociologist's prerogative to identify the social factors which had intervened. Hence, the sociology of science was effectively a sociology of scientists, a series of analyses of relationships between people who just happened to be scientists.[4]

The revision of this view has had fascinating consequences. It turns out to be both necessary and desirable to take the content of science seriously, to attempt to apply sociological analyses of the generation of scientific knowledge without regard for its (perceived) truth status. These claims alone have amounted to major transgressions of previously established disciplinary competences. Whereas the earlier sociology of scientists coexisted more or less peacefully alongside both the internalist history of science and the objectivist philosophy of science, recent claims of the sociology of scientific knowledge have, not surprisingly, caused consternation because they trespass on the domain previously reserved for these other disciplines. Current sociology of science has been called the 'philosophical' or 'epistemologically relevant internalist sociology of science' (Campbell, 1979).[5].

The restricted view of sociological competence corresponds to the pre-Kuhnian view of science. Concomitant with that outdated view is a distinction between the 'technical' (sometimes 'intellectual' or 'cognitive') aspects of science, on the one hand, and peripheral 'social factors' on the other. This distinction was regarded as definitive of the scientific enterprise; 'social' factors were precisely those factors not germane to 'science itself'; the domain of the social was regarded as outside or (at best) peripheral to the actual science. The argument of the new social study of science is not just that as much attention should be paid to the social as to the cognitive (for that merely engenders a kind of analytic 'parallelism' – discussions of the 'connection between social and cognitive factors' – which preserves the distinction intact, nor indeed that the social should supplant the cognitive (since that merely enjoins endless discussions about the appropriate scope of the social and cognitive – witness the age-old debate between internalism and externalism). Instead, the distinctions themselves need to be transcended. We need to recognise such distinctions as the achievement of science, as a resource for the characterisation of behaviours and practices, and as deeply ingrained in a discourse which sustains its own practice as 'scientific'.

The repeated avowal of a distinction between 'science' and 'social' sustains the image of science as an essentially non-social activity. Thus, Boden's implicit analogy with arguments for social responsibility in the natural sciences presumes the same limited sense of relationship between the context of the generation of AI (discovery of scientific knowledge) and the context of its subsequent use (justification of scientific knowledge). The view that sociology is competent to deal only with the latter is predicated upon the very distinction between 'the scientific' and 'the social' challenged by the recent social study of science. Sociological studies which focus solely on the impact of AI research, to the exclusion of the research activity itself, similarly underwrite the distinction between the scientific and the social. In order to escape the idea that the 'scientific' character of AI research makes sociological investigation inappropriate, we need to rid ourselves of a distinction which unnecessarily restricts the realm of the 'social'.

We have seen how one particular feature of discussions about AI can be explained by drawing upon recent developments in the sociology of scientific knowledge. The distinction between 'the social' and 'the scientific' is a major barrier to a thorough-going sociological analysis of AI. It is similarly possible to use ideas in the sociology of scientific knowledge to explain the nature of discussions about the state of AI research. As an indication of the character of discussions about AI research, we turn to examples taken from the literature on 'expert systems'.

Progress and variation: discussions of expert systems

Work on 'expert systems' is one subfield of AI. Expert systems are entities which perform tasks deemed intelligent, in virtue of their access to and use of a knowledge base. In the U.K. in particular, the term 'intelligent knowledge based systems' (IKBS) is used to speak about the performance of tasks requiring knowledge by either (or both) machines and people. In this uage, 'expert system' is the term for an IKBS designed to make decisions about a specific knowledge domain (Murray and Richardson, 1983).

The notion of the expert system arose out of a more general interest in problem solving. Cognitive psychology has long been concerned to specify 'mechanisms' whereby people solve problems – a classic example is the missionary-cannibal problem.[6] AI has been concerned with the possibilities for designing artificial systems which can also solve them. It is assumed that the capacity for problem-solving is unqiue to humans. Notwithstanding the achievements of certain celebrity animals – dogs like Lassie, and apes like the 'speaking' Nim Chimpsky – humans are claimed to have a far greater capacity for the solution of these problems. Significantly, far less is made of the observation that the *generation* of this particular genre of problems is a uniquely human trait, and probably a culturally and historically specific one at that. As far as we know, the missionary-cannibal problem is not a major concern of dogs and apes.

The kinds of puzzles, games and problems typically used for these investigations are now thought to be largely unlike the kinds of problem which rely upon the solver's access to and use of a knowledge base; most 'brain teasers' are said to be 'knowledge poor' (Hunt, 1982: 260). This has led to the task of finding how solutions are determined for problems normally thought to require human specialists. In practice, expert systems are computer programmes intended to serve as consultants for decision

making. For example, MYCIN is an expert system developed at Stanford University in the mid-1970's in order to assist doctors' selection of antibiotics for patients with severe infections. The expert system comprises two parts. One is the knowledge base containing the facts and heuristics of a particular discipline – blood infections in the case of MYCIN. The second is an inference procedure, a set of rules for the manipulation of the knowledge base.

Discussions of the state of expert systems research exhibit a marked rhetoric of progress. Work on expert systems is said to be at the cutting edge of AI research. Early attempts to produce a relatively small number of powerful techniques for generating intelligent behaviour have gradually waned (Duda and Shortliffe, 1983: 261). 'Problem independent heuristic methods' are now regarded as incapable of wide application. Much better chances of success are said to be assured when the problem to be solved involves more precisely stated 'problems', 'facts' and 'axioms'. For example, the very general facility of language understanding is now thought to be much more difficult to emulate than the deduction of advice based on knowledge about a very narrow class of problems. AI has thus turned from the search for a few powerful all-encompassing techniques of intelligence to a strategy for developing knowledge-based approaches to problem solution. This change in direction is described by practitioners as a 'shift in paradigm' (Goldstein and Papert, 1977). Ironically, the prospects for creating problem solving systems are now thought better in the area of highly esoteric, advanced expertise, than in the realm of common sense deductions; it is easier for machines to solve highly abstract problems than to behave common-sensically.

Expert systems have been widely acclaimed as the applied end of AI research, the long-awaited tangible outcome of research investment. 'The science of artificial intelligence ... is at last emerging from academic obscurity' (Evanczuk and Manuel, 1983: 139). 'Commercial products begin to emerge from decades of research ... expert systems ... herald what could be a new tidal wave' (Manuel and Evanczuk, 1983: 127). AI has come of age and has 'risen above its ancient (sic) image' (Yasaki, 1980: 48). 'One could imagine some use for expert systems in just about any sphere of business, engineering or research' (Webster and Miner, 1982: 60). Expert systems provide 'practical uses for a useless science' (Alexander, 1982: 1). 'Knowledge-based expert systems come of age' (Duda and Gaschnig, 1981: 1).

The expert systems literature also reveals a marked discrepancy in reports about the state of research in expert systems. The extraordinary optimism of some reports is elsewhere countered by considerable caution and pessimism about the achievements to date. On the one hand, expert systems is generally regarded as one of the most active and exciting areas of AI research. On the other, there is considerable concern about the fact that the field currently faces 'fundamental problems' (Davis, 1982; Duda and Shortliffe, 1983). In terms of the number of expert systems in existence, we find claims that 'nearly fifty' had been built by early 1982 (Resnick, Port and Hall, 1982; Manuel and Evanczuk, 1983: 128). In the less popular press, however, it is reported that despite impressive performances by some of these systems, only four of the best known systems are in regular use (Duda and Shortliffe, 1983: 265). According to interviewees at M.I.T. and Stanford University, even this is an overestimate (for example, Szolovits, 1983)!

Progress and variation: the example of the Turing test.

What accounts for the extraordinary rhetoric of progress and the marked variation in descriptions of the state of the art? Although these particular discussions about expert systems are not necessarily representative of discussions about AI as a whole,[7] it is nonetheless worth considering that there are basic features of the AI enterprise which give rise to the rhetoric of progress and to variation in assessments of the state of the art. In order to address this possibility, we examine discussions of the most celebrated suggestion for deciding what counts as 'intelligence' in machines: the Turing test.

The test proposed by Turing (1950) to determine whether or not a machine can think is a variation on the 'imitation game'. A machine and a person are in separate rooms. Both are interrogated by a third party via some sort of teletype set-up. The machine 'passes' the test if the interrogator is unable to determine the difference between the machine and the person.[8] The Turing test is problematic, it is claimed in recent philosophical discussions, because it confuses mere signs of intelligence with what intelligence actually comprises (for example, Block, 1981). Thus, it is said, the Turing test fails to allow for the possibility that although a machine's performance *appears* intelligent, this in itself does not determine whether or not the machine is *actually* intelligent. Indeed suggestions have been made for devices which pass the Turing test but which are 'manifestly not intelligent' (Block, 1981; also Colby, 1973; Heiser et al, 1980; but see Weizenbaum, 1974, 1976). This argument thus reflects a basic tenet of mundane reasoning: it is possible to treat superficial appearances as not necessarily representative of the underlying reality from which they originate. We can interpret a particular action as stupid, even while maintaining the possibility (belief, knowledge, conviction or whatever) that the actor is in fact bright (clever, intelligent and so on). (For example, 'He wrote an absurd paper, but we know he is capable of better things'). Our apprehension of surface signs does not determine our views about the underlying reality, even though on occasion we may treat them in this manner. What the signs seem to point to is always revisable in the light of further signs. In this sense, the philosophical complaint about the Turing test indexes our practical subscription to Quine's (1960, 1969) thesis of the underdetermination of theory by observation.

The difficulty comes when the distinction between surface signs and underlying reality is expressed as a fundamental and inviolable principle of scientific procedure. In this view the distinction between 'what it is to be intelligent' and 'how we tell something is intelligent' corresponds to a distinction between the metaphysical essence of an entity (what it is to be an X) and its epistemological apprehension (how we know something is an X). According to this view, the mistake of Turing (and people like him) is illegitimately to confuse the metaphysical with the epistemological.

In order to appreciate the consequences of this philosophical distinction, we need to recall a second feature of the post-Kuhnian shift in attitudes to science. Although many sociologists have now abandoned the idea that there is necessarily anything privileged about science, many have yet to appreciate the full significance of the point that the very notion of 'science' has an interpretive flexibility. In order to argue that science is an ordinary social phenomenon, sociologists first organise observations (of action and practices) and claim their specificity to a distinct cultural entity; they label this entity 'science' or 'scientific practice'. However, it is clear that constituting the object of inquiry in this way involves a *de facto* demarcation. Materials are organised

in such a way that sociological investigation yields findings 'about science'. Unfortunately, this procedure overlooks the fact that the initial construal of a distinct object for study is no less the upshot of adopting the discourse of science than granting epistemological privilege to its 'scientific' (as opposed to 'social') aspects. The very notion of 'science' is as much a flexible interpretive resource as the distinction beween the 'cognitive' and the 'social'.

Two examples will elucidate this point. First, the experience of the recent ethnography of science suggests the adage: 'the science is always elsewhere'. In several attempts to negotiate participant observation in scientific laboratories, practitioners repeatedly told me that the work of their particular laboratory was not really representative of science.[9] Sometimes the claim that their laboratory was not representative was made in disciplinary terms (for example, the claim that neuroendocrinology is hardly representative of science; it is after all merely a soft science; the real hard work goes on in physics); sometimes it took the form of a claim about differences between research areas (for example, that observing work in a solid state physics laboratory wasn't the best place to get a handle on science; for this I should study people working in high energy particle physics). The scientists making these arguments may have had many reasons for wishing to dissuade me from studying their particular laboratory. (In neither of the above two cases were they successful!) But the point brought home to me was that the very notion of 'where the science is' has an interpretive flexibility which enables practitioners to formulate the presence (and defining characteristics) of science in a variety of different ways (cf. Moerman, 1965, 1968). What they construed as the defining characteristic of science, 'what science actually is', was used as a resource in attempting to deter my request for access. Evidently, 'science' is as much an interpretive resource for natives as for ethnographers (Sharrock and Anderson, 1982).

Second, the traditional versions of Science handed down by generations of objectivist philosophy are notably at odds with the descriptions of mundane, day-to-day laboratory work in recent ethnographic studies of science (Knorr-Cetina, 1981; Latour and Woolgar, 1979). Since science at the laboratory bench appears little different from everyday practical reasoning, we are again led to ask: where is the science? Since the science vanishes from our grasp just when we think we have reached the heart of the matter, we are left with the rather weak argument that what we observed must have been scientific since it took place within a laboratory.

Once again, we see the illusory character of Science. In line with recent calls for an analysis of scientific discourse (for example, Gilbert and Mulkay, 1984), my suggestion is that science exists only in and through occasioned appeals to certain ideals of procedure, rules and moral imperatives. Scientists deploy a repertoire of devices for evaluating and appealing to the scientific character of what they are doing.

Just as the science is always elsewhere, and hence scientists tell me that my sociological investigation will never capture its content (certainly not in *their* laboratory, anyway), so too is the intelligence always elsewhere in the philosophical distinction between 'what intelligence actually is' and 'mere signs of intelligence' as detected by the Turing Test. This distinction is treated as a given, pre-theoretic duality, against which any claim to empirical detection of intelligence can be assessed.

This distinction has two important consequences. First, the claim that the metaphysical entity is not reducible to epistemology helps establish and preserve the object of inquiry. By holding that the underlying entity may be different from the

indication of surface signs, one reaffirms the existence of the object. Although we may never have it in hand, the object's existence is assured. From time to time it may reveal its true character, but there seems no way of recognising when this is happening. Second, the postulation of an object's wholly metaphysical character relieves it of the responsibility of revealing itself. Thus it can always be said, in virtue of the distinction between surface signs and underlying reality, that we have not *yet* got the object. In the absence of any stipulated mechanisms for closure, the distinction evinced by philosophers like Block perpetuates the task of the science indefinitely. The object recedes into the distance, taking refuge in a more remote corner of metaphysical space, and thereby occasioning yet further efforts by its hapless pursuers. The philosophical distinction between the metaphysical and the epistemological thus provides the research enterprise with a powerful dynamic. In a similar way, the argument that any specific appearance of 'intelligence' may not turn out actually to be 'intelligence' provides the AI community with a seemingly endless research programme. The latest manifestation of intelligent performance merely occasions the redefinition of what (after all) intelligence comprises.[10]

We can illustrate this point with the simple example of an imaginary device which detects the onset of advertisements, commercial messages and other nuisance interruptions during television programmes.[11] This would permit the television to be turned off (or at least muted) during these interruptions. Two quite different reactions to the operation of the device are possible. We might be completely satisfied with its efficient execution of the desired task. We might speak of the device 'knowing' when to spare us from the misery of 'messages from our sponsor'. Alternatively, we might be disappointed to discover that the device worked 'merely' by detecting some change in the electronic signal at the onset of commercial breaks. Its operation, on this view, might be said to be 'entirely mechanical' and not 'really' what we would call 'intelligent'. We might say it was 'unable to determine changes in story line', that it 'failed to see' that the commercial was substantively different from the interrupted programme. Thus, although on one level we could be perfectly happy with its 'intelligent operation', we could also argue that the device was 'not really intelligent'. Importantly, the latter view redefines and thus reserves the attribute of 'intelligence' for some future assessment of performance. The way is thereby cleared for further research into devices which are 'really intelligent', where this is (temporarily) equated with the capacity to analyse story lines, content, tenor of presentation and so on.[12]

The preceding analysis reveals three key features of discussions about AI. Firstly, pre-Kuhnian notions of the scientific/technical character of AI research endorse distinctions between the realms of 'scientific' and 'social' which withhold all but peripheral phenomena from sociological purview. Secondly, the interpretive flexibility associated with the concepts of 'expertise' and 'intelligence' facilitates the continual redefinition of the operational correlate of these entities. Thirdly, the classic distinction between surface appearance and underlying essence, in this case the distinction between 'what intelligence actually is' and 'mere signs of intelligence', provides an important dynamic for the research enterprise.

These three features begin to account for the rhetoric of progress which characterises discussions (and justifications) of the field and for the variation in reports of the state of AI research. Discrepancies in reports of the state of expert systems research could also be explained by noting that different versions are likely in different reporting contexts. We might expect optimistic representations of the vitality, achievements and potential

of the field from those involved in marketing expert systems. But whatever the interests of the marketing entrepreneurs, our analysis suggests that an endemic feature of AI discourse facilitates the differential portrayal of the research product. The interpretive flexibility of the object of the enterprise permits variations in the definition of 'expert systems'. What counts as 'expertise' is at the heart of these discrepancies. Not only does it have different meanings and significance for different audiences (on different occasions, for different interests and so on). The continual definition and redefinition of 'expertise' draws on and sustains the inner dynamic of the research enterprise. This dynamic is crucial to the rhetoric of progress which accompanies portrayals of work in expert systems and in AI more generally.

These features of AI discourse also begin to account for the extraordinary marketability of the future products of AI research. The promise of future achievement hinges on the flexibility which makes redefinition possible. Alexander (1982: 2) reports a saying among AI researchers that 'If it's useful it isn't AI'. The successful design of a machine which *appears* to work intelligently can be taken as the grounds for presuming that intelligence is, after all, something more than the machine manifests. Hence the need to build a machine which is 'really intelligent'. The general feeling in the AI community that the 'Turing Test is now too easy' (Shurkin, 1983: 73) similarly indexes a redefinition of what counts as intelligence. The feeling that AI research has encountered severe problems is perhaps best understood as a reflection of the move away from designing (mere) classification systems (such as MYCIN's medical diagnosis consultation) to building expert systems for the solution of synthetic problems, such as planning and the *de novo* solution of problems. Similarly, the current claim in the literature that expert systems are 'most successful' when they are amenable to friendly interrogation reflects a recent expansion of the definition of expertise to include the ability to display the grounds for their reasoning.

Implications

What are the implications of the features of AI discourse outlined above for attempts to understand the phenomenon sociologically? Clearly, our sociological programme will reflect the extent to which we adopt the distinctions, concepts and assumptions of AI discourse. I have suggested that the uncritical adoption of the distinction between the 'cognitive' and 'social' restricts the scope of sociology to the impact and context of the use of AI. Similarly, adherence to the view that the phenomenon for AI investigation are the inner processes responsible for 'thought' and 'intelligence', will place these entities beyond the reach of mere observational social science. By being insufficiently alert to the interpretive flexibility of notions of 'intelligence', sociology is left uncertain as to the intelligent character of its subjects and has to wait upon the outcome of what (currently) seems an interminable research 'progression'.

The dangers of uncritically adopting the discourse of the subjects of study are well illustrated by noting that the distinction between man and machine has been used to considerable effect by some practitioners in the expert systems field. By virtue of their 'political' skills (Latour, 1983, 1984), certain individuals have become highly effective salespersons. In particular, they have mobilised the distinction between man and machine in claiming their own particular (human) expertise to speak about expert systems (machines). They thus define the nature and character of the object of study,

they establish that these are indeed the proper objects of investigation and they claim to be uniquely competent in speaking on behalf of these objects. The rest of us are obliged to defer to what these privileged spokesmen have to say about expert systems. In claiming to be especially qualified to define and articulate associations between different expert systems, they establish themselves as experts on the social order of expert systems. They claim to be especially well placed to pronounce upon the relevance of these objects' behaviour for the wider world. More significantly, they claim a particular skill in predicting and explaining the emergence of new and better objects. The inner dynamic of the enterprise ensures the constant search for different ('improved', as they say) species of expert system. The entrepreneurs are experts in marketing the potential of their work, i.e. they have the ability to speak on behalf of an as yet unknown population of future machines. The effectiveness of their skillful deployment of the discourse of AI is perhaps mirrored in our more mundane experiences as neophyte users of personal computers: we 'know' that the very next word processor we buy will soon be obsolete.

In a sense, our uncritical adoption of the man-machine distinction would amount to compliance with the arguments of the entrepreneurs. More importantly, this would overlook the need to develop an appreciation of the generation and use of these distinctions. Instead of taking them on board, we need to understand how such fundamental distinctions underpin the AI phenomena.

What then of the prospects of an association between sociology and AI? In general, the particular style of sociology we espouse will depend on our presumptions about the character of our subjects, the nature of their behaviour and so on. Our willingness to adopt the features of AI discourse identified in this paper will hinge upon our preconceptions about the nature of machines and human behaviour and this, in turn, will shape the relationship between sociology and AI. In particular, our stance with respect to these features of AI discourse will have a direct bearing on whether we construe machines as subjects or objects of sociological analysis.

One option is to develop a sociology of the same phenomena which AI takes as its subject matter. For example, the work is conversational and interaction analysis might provide useful insights for those attempting to construct systems capable of understanding natural language (Gilbert and Heath, 1985). In developing this line of inquiry, we can anticipate two main problems. Firstly, the axiom that activities such as knowing, understanding and so on are fundamentally social is bound to conflict with some current assumptions in hard line AI research. According to writers like Coulter (1983), the assumptions of cognitive theory are just too much at odds with the kind of neo-Wittgensteinian sociology which argues for the socially constituted character of knowledge, expertise, the use of rules and so on (cf. Suchman, 1985). To the extent that AI conflicts with the central arguments of a neo-Wittgensteinian sociology, the prospects of a useful sociology *for* AI seem bleak. Secondly, the utility of a sociological analysis of 'cognitive' activities for the AI project will presumably depend on the extent to which sociological results are formalisable (and hence programmable). Whether or not adherents to the neo-Wittgensteinian line are willing to allow the codifiability of the results of their analyses is an open question (see Woolgar, forthcoming). Clearly, an increased interaction between AI and sociology may lead to some modification of assumptions by either party. If, for example, conversational analysis turns out to be (that is, becomes defined as) 'useful' to the development of systems for natural language understanding, this will provide an

interesting comment on the sense in which conversational analysis can be said to generate 'actual results'.

An alternative (but not necesssarily incompatible) option is to develop a sociology *of* AI research practice. This should proceed so as to maximise our commitment to the social character of those activities designated 'cognitive', but it should eschew adoption of the analytic premisses built into AI discourse. Let us briefly review four different styles of a sociology of AI available to us.

Firstly, we could imagine a sociology of AI researchers, but this, by analogy with the distinction between a sociology of scientists and a sociology of science, might tell us little about the products of their research. We would be looking at the researchers rather than the research practice. Secondly, we can adopt the more current sociology of science position that the products of AI research are socially constructed. Under this rubric one would develop a sociology of the characterisation, design and use of intelligent machines; the machines would be portrayed as socially constituted objects. Note, however, that this approach grants priority to humans as constructing agents, and this implicitly adopts the key distinction between humans and machines which pervades AI discourse.[13] A third kind of sociology of AI would construe intelligent machines as the subjects of study. There seem no difficulties of principle in using standard sociological methods in this approach. We can imagine the successful programming of a machine to produce responses to a questionnaire or interview questions. And presumably a participant observation study would entail no more than successful 'man-machine interaction' This project will only strike us as bizarre to the extent that we are unwilling to grant human intelligence to intelligent machines.[14] Yet the grounds for restricting our attention to the activities of machines alone seem just as arbitrary as confining our study to the practice of humans. Both the second and third alternatives involve the implicit adoption of the human-machine distinction.

Conclusion

The tentative conclusion of this paper, then, is that we adopt a fourth alternative style. Our sociology of machine intelligence has to do more than merely adopt the discourse of AI. It should instead take as topic the dichotomies and distinctions which characterise and sustain this discourse. To achieve this, we need to eschew approaches which are unnecessarily parasitic on participants' dichotomies, and develop a sociological approach which takes as its focus the human/mechanical language community; the community composed of 'expert machines and machine experts'. Clearly, this would entail an empirical investigation which goes beyond the largely conceptual analysis outlined in this paper. For example, we need to investigate the relationship between the pronouncements of spokesmen on behalf of AI[15] and the practical day-to-day activities of AI researchers. What circumstances generate these public accounts of the importance of AI, and how do its proponents respond to the argument that the achievements of AI should not be evaluated in terms of their relevance for 'intelligence' or any other 'mental' phenomena?[16]

The AI phenomenon has a strategic importance for sociology in at least two senses. Firstly, AI is a technology which provides an interesting test case for attempts to extend approaches in the sociology of scientific knowledge to the phenomenon of machines more generally (cf. Pinch and Bijker et al, forthcoming). Of course, much of

the AI (and expert systems) research discussed in this paper comprises work on a particular class of algorithms which just happen to be tested on computers. Nonetheless the use of the term 'machine' to describe the workings of these algorithms raises the possibility of extending this same kind of approach to other classes of machine. To what extent can we develop a sociological study of the human/mechanical language community where the 'machines' in question are, say, bicycles, missiles or food processors? Secondly, the AI phenomenon provides an important occasion for reassessing one of the basic axioms of sociology, viz. the claim that there is a sense in which human behaviour can be understood as distinctively 'social'. More generally, perhaps, AI provides the opportunity for reevaluating our preconceptions about behaviour, action, its origins and agency, and, most significantly, our attempts to understand. I suggest it is instructive to press closely the claim that there is something special about human behaviour. Or, to put it more carefully (since it is not my intention to legislate on the 'actual character' of human behaviour), it is important to examine how sociology presumes human behaviour to be unlike the performance of a machine. How do prevalent conceptions of machine activity and social behaviour shape sociological explanation? Given the reductionist tendencies of many sociological styles, the implicit assumption that virtually any activity can be analysed sociologically, it is interesting to ask why sociology should stop short when it comes to machines. How exactly do presumptions about the 'social' exclude machine-like activity from the purview of sociological investigation? Why not a sociology of machines? Are artificially intelligent machines sufficiently like humans to be treated as the subjects of sociological inquiry? Or, to reverse the more usual query, in what sense can we continue to presume that human intelligence is not artificial?

Hitherto abstract concerns in the philosophy of the social sciences can now be broached empirically by reference to the recent attempts of AI researchers to probe the limit of the distinction between human behaviour and machine activity. Thus the question of whether there are essential differences between humans and machines can be addressed with respect to attempts to develop a sub-class of machines which are, arguably, endowed with a human capability, intelligence. The phenomenon of AI provides an opportunity for investigating how presumptions of the distinction between human and machine delimit social inquiry.

Notes

1 The revision of earlier versions of the argument here has benefitted substantially from comments by Jeff Coulter, James Fleck, Gonzalo Munevar, Lucy Suchman, Anna Wynne, anonymous referees and participants in the workshop on 'Discourse and Reflexivity', Department of Sociology, Brunel University, 31st March – 1st April 1984.

2 In her well known textbook, Margaret Boden argues that (AI) 'offers an illuminating theoretical metaphor for the mind that allows psychological questions to be posed with greater clarity than before' (1977: 473). We can similarly anticipate benefits for philosophical investigations into the nature of mind, since the phenomenon of AI provides an impetus for attempts better to articulate what mind consists of, what can be legitimately described as 'intelligence' and so on. Almost all introductory sections of contributions to the AI literature include a ritual citation of the relevance of AI for philosophy, psychology (for example, Haugeland, 1981), linguistics and even physics (for example, Gregory, 1981). Of course, many other commentators have offered rather different visions of the import of artifical intelligence and many have strongly contested its potential (for example, Dreyfus, 1979; Coulter, 1983; Weizenbaum, 1976).

3 Some exceptions are Gilbert and Heath (1985), Fleck (1982, 1984).

4 As a methodological corollary, it was sufficient to rely upon the formal writings of science, the memoirs and recollections of (usually eminent) scientists, their disengaged interview responses and whatever 'compliance documents' (Garfinkel et al, 1981: 133) the sociologist could persuade them to complete.

5 A methodological corollary of the shift in epistemological preconceptions is that it is now possible to undertake detailed participant observation studies of the practice of science. The fact that the culture of the laboratory is scientific is no reason for the exclusion of the sociologist. See, for example, Latour and Woolgar, 1979; Knorr-Cetina, 1981; Lynch, 1985; Traweek, forthcoming.

6 For an exposition of the missionary-cannibal problem, and solutions, see, for example, Hunt (1982: 25 and 365).

7 Some practitioners would argue that expert systems work does not properly count as AI because it is applied. In addition, as is discussed below, certain special considerations apply to expert systems work in view of the marketability of its products.

8 For an introductory discussion of the subleties of the test see, for example, Hofstadter (1981)

9 Of course, this was partly due to expressing my interest 'in science' and 'in the way science works in practice' rather than, say, 'in the technical skills involved in the work of your particular laboratory'.

10 The empirically interesting question which follows is that of how any particular instance of scientific inquiry ends. How do practitioners conclude that they now have what they were after and, most importantly, what are the associated changes in discussions of the connection between the metaphysical and the epistemological?

11 Since first writing this, I have been told that such a device indeed exists. It has been developed in the U.S.A. by a well known electronics company, in order to enable editing of video taped TV programmes. But it is not being marketed, so the the story goes, due to enormous pressure from the TV networks.

12 Two distinct kinds of criteria are being used here (Suchman, 1984). The example implies that we might be willing to attribute intelligence on the basis of effect (performance) alone, but unwilling to do so on the basis of the way it operates (mechanism). This distinction mirrors the difference between those AI researchers committed to the simulation of cognitive mechanisms and those who believe that the prime task is to mimic human behaviour (performance) by whatever means this is achieved. One of several difficulties, however, is that these criteria are not unproblematically distinct. For example, a description of a machine's performance may well involve an assessment of the means by which it performs. Similarly, we can see from this example that the assessment of intelligence in terms of mechanism projects a further ambiguous assessment of performance viz. 'the capacity to analyse story lines etc.' This suggests that an important topic for investigation is the ways in which descriptions of machine activity are construed as descriptions of, say, performance rather than mechanism.

13 Although this paper has begun to reveal some sources of restriction on a sociology of machine intelligence, our discussion has concentrated on the strategic practices of members of the AI community, and on the discussions and arguments between others who speak about AI. In identifying aspects of the discourse on AI we ourselves have unwittingly adopted one of its key features: the distinction between humans and machines. Thus we have taken humans, rather than machines, as the subject of discussion and our analysis of AI discourse has been an analysis of human discourse on AI.

14 Precisely because we use a metaphor of human communication to speak of systems of interacting intelligent machines, it is less difficult to conceive of a network study of machine communication.

15 Obviously, we should be wary of saddling philosophers like Block with an ontological commitment to the distinction between the metaphysical and the epistemological. These are, after all, occasioned pronouncements. Nonetheless, they do speak on behalf of AI research.

Does the distinction championed by writers like Block, the difference between metaphysical objects and their epistemological correlates, have any currency in the actual practice of AI research? One version is that AI researchers simply want to make their machines perform a task; on this view 'talking INTELLIGENCE', by analogy with the activity of 'talking SCIENCE', occurs mainly in peripheral and programmatic discussions of AI, and such talk has no bearing on the practice of AI. Whether or not AI practitioners talk INTELLIGENCE in the course of their practice is an open question. If they are anything like the laboratory scientists we already know about, their shop floor practice, their talk in the course of evaluating programme output, for example, is unlikely to articulate the philosopher's distinction. Nonetheless, it would be interesting to discover that there was no reference to INTELLIGENCE in the course of AI practice, because then the enterprise would rely entirely for justification of its greater significance on outsiders who themselves did not contribute to the practice.

16 Coulter (1983) argues that cognitive theory depends on a false equivalence between the achievements of AI and the operation of cognitive processes; that arguments for establishing a phenomenon (the mind) for scientific investigation are spurious and the notion of 'cognitive processes operating at the unconscious level' is fallacious. There is no reasonable basis to suppose such phenomena exist, according to Coulter, except through the illegitimate bludgeoning and reformulation of ordinary concepts of human behaviour and action. Arguing for the removal of 'unconscious cognitive processes' as the appropriate topic of investigation does not imply a criticism of the technical achievements of AI. It is simply that the assessment of such achievements in terms of their relevance for mysterious (and artificially construed) 'cognitive processes' is inappropriate: 'I do not think that (AI) need be assessed *at all* in connection with psycho-physiological meta-theory, any more than progress in advanced cartography needs to be assessed in terms of the need for theorising about children's (or non-technical adults') map-using capacities.' (Coulter, 1983: 25, emphasis in original).

References

ALEXANDER, T. (1982) 'Practical uses for a "useless" science' *Fortune* (May 31).

BIJKER, W.E., HUGHES, T. and PINCH, T.J. (eds.) (forthcoming) *New Directions in the Social Study of Technology*

BLOCK, N. (1981) 'Psychologism and behaviourism'. *The Philosophical Review.* 90 no. 1 (January): 5-43.

BODEN, M. (1977) *Artificial Intelligence and Natural Man* (Sussex: Harvester).

CAMPBELL, D. (1979) *Descriptive Epistemology* (Cambridge, Mass.: Harvard University Press).

COLBY, K.M. (1973) 'Simulation of Belief Systems' *in* Schank and Colby (eds.)

COULTER, J. (1983) *Rethinking Cognitive Theory* (London: Macmillan).

DAVIS, R. (1982) 'Expert systems: where are we and where do we go from here?' A.I. Laboratory Memo No. 665, MIT, June.

DREYFUS, H. (1979) *What Computers Can't Do* (New York: Basic Books, 2nd Edn.).

DUDA, R.O. and GASCHNIG, J.G. (1981) 'Knowledge-based expert systems come of age'. *Byte* (September): 238-281.

DUDA, R.O. and SHORTLIFFE, E.H. (1983) 'Expert systems research'. *Science* 220: 261-268.

EVANCZUK, S. and MANUEL, T. (1983) 'Practical systems use natural languages and store human expertise' *Electronics* (December 1): 139-145.

FEIGENBAUM, E.A. and MCCORDUCK, P. (1983) *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge To The World* (London: Addison-Wesley).

FLECK, J. (1982) 'Development and establishment in artificial intelligence' in Elias, N., Martins, H. and Whitley, R. (eds.) *Scientific Establishments and Hierarchies. Sociology of the Sciences Yearbook* 6: 169-217. (Dordrecht: Reidel).

FLECK, J. (1984) 'Artificial Intelligence and Industrial Robots: An Automatic End For Utopian Thought?' *in* Mendelsohn, E. and Nowotny, H. (eds.) *Nineteen Eighty-Four:Science Between Utopia and Dystopia. Sociology of the Sciences Yearbook* vol. 8: 189-231 (Dordrecht: Reidel).
GARFINKEL, H., LYNCH, M. and LIVINGSTON, E. (1981) 'The work of a discovering science construed with materials from the optically discovered pulsar'. *Philosophy of the Social Sciences* 11: 131-158.
GEERTZ, C. (1973) *The Interpretation of Cultures* (New York: Basic Books)
GIERYN, T.F. (1983) 'Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists'. *American Sociological Review* 48: 781-795.
GILBERT, G.N. and HEATH, C. (eds.) (1985) *Social Action and Artificial Intelligence* (Aldershot: Gower Press).
GILBERT, G.N. and MULKAY, M.J. (1984) *Opening Pandora's Box: A Sociological Analysis Of Scientists' Discourse* (Cambridge: University Press).
GOLDSTEIN, I. and PAPERT, S. (1977) *Cognitive Science* 1:84
GREGORY, R. (1981) *Mind In Science: A History Of Explanations In Psychology And Physics* (Cambridge: University Press).
HAUGELAND, J. (ed.) (1981) *Mind Design: Philosophy, Psychology, Artificial Intelligence* (Cambridge, Mass.: MIT Press).
HEISER, J.F., COLBY, K.M., FAUGHT, W.S. and PARKINSON, K.C. (1980) 'Can Psychiatrists Distinguish A Computer Simulation From The Real Thing?' *Journal Of Psychiatric Research* 15: 149-162.
HOFSTADTER, D.R. (1981) 'Metamagical themas: a coffeehouse conversation on the Turing test to determine if a machine can think'. *Scientific American* May: 15-36.
HUNT, M. (1982) *The Universe Within: a new science explores the human mind* (New York: Simon and Schuster)
KNORR-CETINA, K.D. (1981) *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*
LATOUR, B. (1983) 'Give Me A Laboratory And I Will Raise The World' *in* Knorr-Cetina, K.D. and Mulkay, M.J. (eds.) *Science Observed: Perspectives On The Social Study of Science* pp.141-70. (London:Sage).
LATOUR, B. (1984) *Les Microbes: Guerre et Paix et Irréductions* (Paris: Pandore).
LATOUR, B. and WOOLGAR, S. (1979) *Laboratory Life: The Social Construction Of Scientific Facts* (Beverly Hills: Sage).
LYNCH, M. (1985) *Art and Artifact In Laboratory Science: A Study Of Shop Work and Shop Talk In A Research Laboratory* (London: Routledge and Kegan Paul).
MANUEL, T. and EVANCZUK, S. (1983) 'Commercial products begin to emerge from decades of research'. *Electronics* (November 3): 127-137.
MOERMAN, M. (1965) 'Who Are The Lue?' *American Anthropologist* 67: 1215-30.
MOERMAN, M. (1968) 'Being Lue: the use and abuse of ethnic identification'. *in* Helm, J. (ed.) *Essays On The Problem Of The Tribe* (University of Washington Press)
MULKAY, M.J. (1979) *Science and the Sociology of Knowledge* (London: Allen and Unwin).
MURRAY, L.A. and RICHARDSON, J.T. (1983) 'The contribution of the social sciences to the development and understanding of intelligent knowledge-based systems'. Report prepared for the Education and Human Development Committee of the Social Science Research Council, November.
PINCH, T.J. and BIJKER, W.E. (1984) 'The Social Construction of Facts and Artefacts or How The Sociology of Science and the Sociology of Technology Might Benefit Each Other'. *Social Studies of Science* 14: 399-441.
PRATT, V. (1978) *The Philosophy of the Social Sciences* (London: Methuen)
QUINE, W.v.O. (1960) *Word and Object* (New York: Wiley)
QUINE, W.v.O. (1969) *Ontological Relativity and Other Essays* (London: Columbia University Press)

RESNICK, M., PORT, O. and HALL, A. (1982) 'Artificial intelligence: the second computer age begins'. *Business Week* (March 8): 66-75
SCHANK, R.C. and COLBY, K.M. (eds.) (1973) *Computer Models of Thought and Language* (San Francisco, Freeman)
SHARROCK, W.W. and ANDERSON, R.J. (1982) 'On the demise of the native: some observations on and a proposal for ethnography'. *Human Studies* 5: 119-136.
SHURKIN, J.N. (1983) 'Expert systems: the practical face of artificial intelligence'. *Technology Review* (November/December): 72-78.
SUCHMAN, L. (1984) Private communication, May 17
SUCHMAN, L. (1985) *Plans and Situated Actions: the problem of human-machine communication* (Palo Alto, California: Xerox Corporation ISL-6)
SZOLOVITS, P. (1983) Interview, M.I.T., 2nd November
TRAWEEK, S. (forthcoming) *Uptime, Downtime, Spacetime and Power: an Ethnology of the Particle Physics Community in Japan and the United States.*
TURING, A. (1950) 'Computing machinery and intelligence'. *Mind* 59: 433-460
TURKLE, A. (1982) 'The subjective computer: a study in the psychology of personal computation'. *Social Studies of Science* 12: 173-205.
TURKLE, S. (1984) *The Second Self: The Human Spirit in a Computer Culture* (New York: Simon and Schuster)
WEBSTER, R. and MINER, L. (1982) 'Expert systems: programming problem-solving' *Technology* 2 (January/February): 62-73
WEIZENBAUM, J. (1976) *Computer Power and Human Reason: From Judgement to Calculation,* San Francisco: Freeman.
WOOLGAR, S. (forthcoming) 'Reconstructing man and machine: a note on sociological critiques of cognitivism'. *in* Bijker, W., Hughes, T. and Pinch, T. (eds.) *New Directions in Social Studies of Technology.*
YASAKI, E.K. (1980) 'AI comes of age'. *Datamation* (October): 48-54

Biographical Note: STEVE WOOLGAR BA (1972) PhD (1978; Emmanuel College, Cambridge) is lecturer in sociology and social anthropology at Brunel University. He is co-author (with Bruno Latour) of *Laboratory Life: the social construction of scientific facts* (Sage, 1979).

Address: Department of Sociology, Brunel University, Uxbridge, Middlesex UB8 3PH.