ARTIFICIAL KNOWING

Gender and the Thinking Machine

Alison Adam

London and New York

First published 1998
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Phototypeset in Garamond by Intype London Ltd
Printed and bound in Great Britain by
Creative Print and Design (Wales), Ebbw Vale

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
Adam, Alison,
Artificial Knowing: Gender and the thinking machine/Alison Adam.
Includes bibliographical references
1. Women–Effect of technological innovations on. 2. Artificial intelligence–Social aspects. 3. Expert systems (Computer science) 4. Knowledge representation (Information theory) 5. Feminist theory. I. Title.
HQ1233.A33 1997
306.4'6—dc21 97–16346

ISBN 0–415–12962–1 (hbk)
ISBN 0–415–12963–x (pbk)

INTRODUCTION

This book is about the ways in which gender, in the form of concepts of masculinity and femininity, is inscribed, in implicit ways, in a type of computer system which comes under the heading of 'artificial intelligence' or AI. It has proved far from easy to arrive at a suitable one line definition for this project, and particularly, to describe succinctly the relationship between gender and AI. I have chosen the term, 'inscription', as a shorthand for that relationship. In talking of 'inscribing' and 'inscription' I mean something similar, but not identical, to the use of these terms in the writing of Madeleine Akrich (1992). Akrich describes the way in which designers of technological objects inscribe a vision or prediction of the world in a new object, defining a framework of action along with actors and the space in which they act. Similarly I shall argue that a gendered vision of the world is inscribed in the technology of AI, albeit in a subtle way which must be uncovered or 'de-scribed'.

If a one-line definition is difficult enough, then so too is an appropriate title. By the title, 'Artificial Knowing', I am referring to the type of knowing which goes on in the 'thinking machines', i.e. the computer systems, simulations and robots which comprise the technical objects of AI. I intend no special commitment to realism in this. My aim is rather to set up an implicit contrast with what might be taken as the 'real knowing' of human actors, especially the knowing of women, which I shall argue is left out of AI's thinking machines.

Although I look at definitions of what is meant by AI in some detail in chapter two, it is worth giving some indication, at this point, of what is covered by the term artificial intelligence, or AI. AI refers to a class of computer system designed to model some aspect of human intelligence, whether it be learning (machine learning), moving around and interacting in the world (robotics and vision), reasoning towards a solution to a problem (search strategies), using natural language, modelling intelligence according to neural models (neural networks or connectionism) or having expert knowledge of some subject (expert or knowledge-based systems). In talking of AI systems, I am using nothing more than a shorthand to

describe AI computer (or robotics) systems which consist of one or more related programs.

Many AI systems run on personal computers and workstations and some require specialist hardware largely found in academic settings. In both these cases, human interaction takes place mainly through a keyboard, mouse, perhaps a microphone or a similar input device. The 'results' are displayed on the computer screen. Many AI systems operate in this way, from expert systems to the more specialized artificial life simulations of populations which I describe in chapter five, the latter often involving considerable use of screen-based graphics.

Industrial robots have been with us for some time, even to the extent that their use in car manufacture was used as a marketing ploy in the late 1970s and early 1980s. Unlike screen-based systems, but like industrial robots, AI robots have a kind of physical embodiment. They may not look very different from industrial robots, although in fact, visually, they can appear much less sophisticated than their industrial counterparts as they are research objects rather than instruments of industrial production. In contrast to industrial robotics, we would not expect safety, reliability and longevity to be the most major issues right at the forefront of AI robotics research.

In discussing screen-based AI and robotics, my overall concern is both to demystify and to put AI computer systems in context. On the surface they 'look' very much like other types of computer based system. Where they are different, is in their claims to model aspects of human intelligence. It is these claims which I investigate, with respect to gender, in what follows.

Working on an AI research project in the mid-1980s, I found myself becoming concerned with the concepts we were trying to represent and particularly with what sort of knowledge was to be captured in our computer systems. In the original project on which I worked, this involved UK Social Security law. Social Security law is an area which invites controversy as, at bottom, the policy making which goes into making the law is highly normative: in saying how certain categories of people ought to be treated it is also, in a less obvious sense, saying how these people ought to behave. The whole controversy over legal AI systems revolves around the question of whether it is possible, or meaningful, to represent the subtle nuances of legal decision-making in a computer system. The refining, or 'finessing' away of detail, something which I refer to later in the text, is worrying: the classic assumption is that we can somehow identify 'social factors' which can be factored out, leaving a realm of the purely technical underneath. Such a view is challenged by the research of Madeleine Akrich (1992) and Bruno Latour (1992) in relation to technology, in general, and by Steve Woolgar (1985) in respect of computer technology, in particular.

Thinking about all these concerns in relation to feminism suggests that

there are some very problematic things going on. For not only could AI systems be used to promulgate what might be undesirably normative views of women and other groups, they are also to be implicated in the process of refining away the plurality of views which is such an essential part of the feminist project, and in particular the feminist political project. There are, of course, some very telling and very well-known criticisms of AI and these I examine, in some detail, in chapter two. But I have always found it difficult to square a philosophical view of AI with what it was like to work on technical AI projects. I began to think that this was because a number of such critiques do not look in any detail at *real* AI systems, and in this respect they are in danger of knocking down straw people. Importantly for the present study none of these has anything to say about gender, and for that reason I argue they are deficient. I wanted to offer a different view of AI, one that was not necessarily in conflict with these traditional criticisms in a philosophical sense, but one where gender assumes centre stage. This led me to consider how knowledge is represented in some real AI systems, how gender is inscribed and maintained therein, through the process of representing that knowledge and various forms of reasoning. Along the way I have tried both to marshal a number of feminist and other resources, and to draw out a set of implications for current and future research.

I have tried to keep the overlap between chapters to a minimum although I am conscious that there is, of necessity, an element of zig-zagging. Some authors, whose work is of such importance to the whole enterprise, appear in almost every chapter; I am thinking, in particular of Hubert Dreyfus (1979; 1992; 1996), Lucy Suchman (1987) and Harry Collins (1990). I hope I will not irritate my audience by telling them too often what they are just about to read and what they have just read. It is also the case, in a sense, that there is a fair amount of apparently introductory material and that it is chapter three before the 'meat course' arrives. To pursue the dining analogy for a moment, I would rather think of my book as a Chinese banquet, made up of lots of little courses of different flavours, where we may taste as much as we like and may go away feeling full but, I hope, not uncomfortable. In any case I felt it was important to set out the theoretical feminist positions which inform the present work. I also wanted to ensure that I made the subject matter of the technology as clear as possible, without making over-simplified blanket assumptions about the nature of AI, and before moving on to a closer look at feminist arguments in relation to the particular technology of AI.

Chapter one introduces the areas of feminist theory against which I wish to locate my study. I argue that AI is best treated as a part of engineering and hence as a technology. Of course engineering has hardly had a neutral history with respect to gender and this is something that I hope will be clear from what follows later. This means that my starting point is gender

and technology literature. However, although I acknowledge that there is a distinction to be made, I do not want to draw too fixed a line between that area and gender and science research. This is partly because the two areas have clearly fed from and into each other over a long period, and also because I have found a number of studies from gender and science, particularly in relation to feminist epistemology, to be useful in developing my own theoretical position. In seeking to problematize the liberal feminist position in relation to women and computing, I discuss the attractions of other feminist positions in the shape of eco-feminism, standpoint theory and postmodern feminism. I cannot, of course, do justice to all the complexities of contemporary feminist thought in such a short space, but I hope I can introduce enough of a theoretical background to illuminate my reasons for turning to writings on feminist epistemology and its critique of traditional epistemology, particularly Western rationalist science.

Chapter two recounts a brief history of symbolic AI, in other words the part of AI which involves representation in symbolic form rather than modelling the structure of the brain. This is followed by a consideration of relevant philosophical and social science critiques. It has been my concern that a number of commentators produce their critiques of AI with a set of assumptions which do not correspond with AI as it is practised out in the world and I wanted to avoid doing the same by making it clear how I saw the historical development of AI. At the same time, this strategy permits me to define the particular area of AI which is my interest, namely symbolic AI, to give a sense of the way it has developed and to say how concepts have arisen and become embedded in the rationale of AI. I consider connectionism briefly, to argue that it is not qualitatively different from symbolic AI, at least in respect of my analysis. I do not pretend that this is a neutral history, even if that could ever be achieved. Although I do not bring in the full force of the feminist critique at this stage, I cannot resist an occasional snipe in the spirit of Sue Curry Jansen's (1992: 11) 'feminist semiological guerrilla warfare'.

There is such a wealth of material from philosophy and the social sciences that criticizes and comments on AI that I felt it would be wrong not to attempt an investigation of relevant writing and, although it is almost all gender blind, it demanded to be considered in bringing together the different elements of my feminist critique. This is particularly so as there are concepts and ideas from this research which come together to inform my theoretical position alongside the more obvious material from gender and technology and feminist epistemology studies. Some of this material takes what I term a 'monolithic' view of AI, i.e. the assumption that AI is, above all, about building an artificial mind or person, so much so, that I wanted to argue that this is problematic both for philosophical and social science research. Although links into my area of research are difficult to make, there do appear to be some hooks, mainly from research

on phenomenology and the body. In looking at some of the philosophical writings of Daniel Dennett, John Searle and Hubert Dreyfus, and in examining sociological and anthropological studies of AI, a number of key issues emerge in the shape of representation, intentionality, agency and culture. Of these, representation, agency and culture re-emerge when feminist theory is considered in later chapters. Chapter two is a long chapter but I hope that by putting all the philosophical, historical and sociological material together in one place I have provided a suitable context for later chapters.

Chapter three examines the way in which the knowing subject is represented in symbolic AI systems. As I argue that AI is informed by mainstream epistemology, this permits me a brief excursion into traditional epistemology to gain an understanding of how it treats the knowing subject. In particular I want to build up a contrast with the idea of the subject in feminist epistemology, and from this I argue that the question of *responsibility* emerges, a topic which is largely absent in mainstream epistemological writing. I then bring the arguments of feminist epistemology to bear on two examples of large symbolic AI projects, Cyc and Soar.[1] Both these appeal to a 'view from nowhere' (Nagel 1986), a view which assumes that it speaks for some universal yet never articulated subject, nowhere yet everywhere at the same time. However it is clear that, at bottom, the authors of such systems regard themselves as the gold standard of universal subjects. It then becomes important to ask whether middle-class male American university professors speak for everyone. Uncovering the subject in these systems is no easy task. The basis of Soar's reasoning, which set the standard for a great deal of later AI research, rests on a fairly limited set of psychological experiments on technically educated, male, US college students working on a very constrained type of example. I argue that it is problematic to extrapolate from these subjects to make universal statements about the way that everyone reasons in a wide variety of situations.

Cyc and Soar are the example AI systems on which I base a number of my arguments. I place a considerably longer description of their designs and scope in chapters three and four, where I address their relationship to arguments from feminist epistemology in some detail. However, as I refer to them in several places throughout the text, it is appropriate to introduce them here.

Both are large AI projects in the symbolic AI tradition described in chapter two. Both orginate in the USA and involve many 'persons years' of effort. Both projects have been running from at least the mid-1980s, although both have roots in earlier work. Although this short description suggests that there are many similarities between the two, the difference in their design philosophies is elaborated in chapters three and four. Whatever their similarities or differences they are significant flagship projects

for symbolic AI. Much of the perceived success or failure of the whole AI enterprise rests on Cyc's and Soar's success or failure.

Chapter four explores both the type of knowledge and the way that knowledge is represented in symbolic AI systems and how this reflects gendered patterns of rationality. I revisit mainstream epistemology to discuss the way that it emphasizes propositional (knowing that) knowledge to the exclusion of skills (knowing how) knowledge and the way that this is mirrored in symbolic AI. The split between rational and irrational is central to feminist critiques of knowledge. Under a traditional view, rationality is associated with the masculine and the life of the mind and irrationality is associated with the feminine and the body. This leads the argument to a consideration of postmodernism in the critique of the rationality/irrationality dualism. As language is so intimately tied to representation it is no surprise that feminism has focused on the role of language in maintaining women's inferior position.

The continental feminists' critique of rationality rests on language in the construction of knowledge and the maintenance of unequal gender relations. Philosophers of language emphasize the structures of mathematical logic, in other words, they assume that the human brain is functionally equivalent to a computer. This allows AI systems to formalize language in its representational structures. In discussing the elevation of propositional or 'knowing that' knowledge over 'knowing how', I use Vrinda Dalmiya and Elizabeth Alcoff's (1993) concept of epistemic discrimination to argue that knowing how knowledge has historically often been connected with what women know and hence is in danger of being marginalized, particularly by the processes of formalization at work in AI systems. But through their concept of 'gender-specific experiential knowing' Dalmiya and Alcoff show that it is not just a simple question of relating knowing how to women's knowledge. There may be many aspects of knowledge, propositional or otherwise which are not amenable to formal representation.

I call on the first of my two example systems, Cyc, to demonstrate the ways in which it mirrors traditional epistemology's emphasis on propositional knowledge, to the extent that there are types of knowledge, or ways of knowing, which it cannot represent. The knowledge represented in both Cyc and Soar is cast in rationalist form where the rational is a masculine norm following Genevieve Lloyd's (1984) characterization of the *Man of Reason*. This norm is maintained by formal languages. For Soar, I particularly want to consider its emphasis on the AI idea of search and goal seeking, which harks back to Aristotelian notions of goals, and can also be seen in terms of the phallocentric urge to a unitary goal described by postmodernist thought. Trying to cast propositional knowledge in the form of rules involves an infinite regress. I will argue that the only way to deal with this regress and do justice to the types of knowledge which are being ignored or marginalized is to somehow bring the role of the body

back into AI systems in order to ground them in the world, particularly as skills-type or knowing how knowledge rests so fundamentally on bodily actions.

Chapter five looks at the problem I have just described, which is usually termed embodiment, and asks the question of how far knowledge of the world belongs to a purely mental realm separate from the body. Category theory and phenomenology, in the shape of research by George Lakoff (1987) and Mark Johnson (1987), suggests that it cannot be separated. Their views on logicism or objectivism broadly concur with Dreyfus's opinions on phenomenology and can be pressed into service alongside a consideration of feminism and the body. The latter serves to reinforce the discussion of the previous chapter on the association of women's labour with bodily things and irrationality, and men's work with the life of the mind and rationality.

There are, however, newer currents in AI which address the problem of embodiment. 'Artificial life' research involves the study of synthetic systems which are designed to exhibit the characters of natural living systems; populations can be modelled over several generations and so they offer the promise of a demonstration of evolutionary biology. 'Emergent' behaviour is behaviour which takes place at levels higher than individual programs.

Following Dennett, I advocate that we should keep hold of the idea that there is nothing mystical about emergence in artificial life systems, despite the fact that their behaviour may not be predictable. Rather than the goal-seeking, searching behaviour prevalent in many AI systems, artificial life systems, by way of contrast, concentrate on the passing on of 'genetic' information from one artificial generation to another. As a form of artificial life, situated robotics (Brooks 1991) and evolutionary robotics (Wheeler 1996) both hold a certain appeal. This is because these robots, rather than incorporating the traditional planning model which tends to immobilize conventional robots, live, instead, in the world of people, and operate by entwining perception and interaction. Rodney Brooks's robots are embedded or situated in the world and they respond to environmental cues.

Artificial life is strongly tied to socio-biological models which are politically problematic for feminists, as they seem to model the worst part of human societies, in the form of combative, aggressive behaviour. The more promising (i.e. more promising than non-embodied, screen-based alternatives) robotics research produces robots which are physically situated and yet not *culturally* situated. In other words they are not functioning members of a social group and have no shared culture. And their type of embodiment leaves out feminine forms of embodiment such as looking after and caring for other bodies. Social science and feminist research, as in the work of Harry Collins (1990) and Lorraine Code (1993), for

example, suggest that it is the cultural dimension which is necessary for an individual to have knowledge of possibly the most important things about their world.

Chapter six looks at possible futures for an AI influenced by feminist ideals. Two small projects are offered as a beginning of what might be possible although I am aware of the contradictions inherent in tackling these projects at all. Then, in a different vein, I look at the rise of interest in 'cyberculture' for broader ways of thinking about intelligent computer technology and feminism. Cyberculture is a masculine youth culture which once again promises an escape from the body. This is not such a great distance from the vision of some AI scientists. Although Donna Haraway's (1991b) cyborg imagery is appealing, in its promise of transgression of traditional boundaries, it is in danger of becoming lost in a cyberfeminism which denies the feminist political project. It is important to keep hold of the political in whatever future we envision.

In this study the major theme is that AI systems, in taking a traditionally gendered approach to knowledge which reflects the style of mainstream epistemology, incorporate a view of the world which tacitly reflects a norm of masculinity, both in terms of the knower and the known. This leaves out other types of knowing subject and knowledge, particularly that which relates to women's ways of knowing.

My 'day job' in a technical AI project, my attempts to finish a PhD in the history of Victorian science and my increasing interest in feminism, gave me a curious starting point for this study. Computer people are fond of articulating 'methodologies' to describe what they do in designing and constructing computer systems. Social scientists and philosophers tend to think more in terms of theories than methodologies. Reflecting on my way of carrying out this study, I would be reluctant to reconstruct the process with any such definite title as 'methodology'. 'Approach' is perhaps a more suitably modest word.

Nevertheless, I have been surprised how much my approach has been influenced by my earlier work in the history of science, in the sense that much of my task has involved the careful sifting through, pulling together and assimilating of texts into a 'story'. The story in this study does not follow an individual or group of human actors as many histories of science might do. Rather it follows a contemporary history of ideas, but then the history of ideas is also an important part of the history of science. My approach has been consciously philosophical. But in writing a book the author must always worry that a different story could have been told. There is always the feeling that you might have started from somewhere else, or, worse still, that you should be everywhere but the place where you happen to be at the present moment. One problem with a philosophical approach is that it does not necessarily lend itself readily to empirical research; this is the case even for the much more realistically

slanted feminist philosophy on which I draw, although there are notable exceptions (see e.g. Belenky et al. 1997; Goldberger et al. 1996; Stanley and Wise 1993).

This has led me to wonder how far this study would have been different had my approach been different, say, more overtly sociological or anthropological. Quite clearly there would be differences. It is unlikely I would have been able to use the particular AI systems that I draw on here as case studies; more likely I would have been reporting participant observer studies of one or two anonymous AI laboratories, along the lines of Diana Forsythe's (1993a; 1993b) research. In this event the case studies would naturally have taken more of the centre stage.

Given all the resources and access in the world, were I able to observe the Cyc and Soar teams whilst wearing an anthropologist's hat, I am sure I would come away with a very different impression. I am quite clear that the published texts of a project do not tell the same story as participant observer story. For instance, Sarah Willis (1997), in her study of a medical information system, points to the way in which the system builders consciously tidy up their stories to present a 'clean picture' to the world.

There is also the very pragmatic consideration, that even with the best of lead times, published accounts are always, at least in some sense, out of date before the ink is dry on the printed page. So we are always dealing with history, even if it is recent history. Yet even so, at least part of my argument is against a view that there is a way of inspecting an independent real world about which we will all agree, so I do not believe that there is one correct view to be had in any case. If, for some readers, this starts to look like a descent into an unbridled postmodern dystopia where no story is better than another, let me say that I believe that there are good reasons for preferring one account to another, which I hope will become clear in the chapters that follow, and that it is possible to retain both an analytical scepticism and realism at the same time. So this means that I see published work, participant studies and indeed other resources too, as 'texts' that all have to be read and where no one account gives a complete true story; all have an 'aboutness' in relation to the projects they describe and all must be interpreted in constructing an analytical framework.

Although this means that the overall shape of my study could have been different, I am doubtful whether the conclusions I draw would have differed substantially. Part of the reason for feeling at least some level of confidence in this, lies in the nature of the feminist philosophy on which I rely. Contrasting feminist philosophy with its traditional counterparts is a heuristic device which I have not originated; I have borrowed a technique which is used time and time again by feminist philosophers. The appeal of such an approach is that it points up the contrast between the more realistic examples which are used in feminist epistemology, such as in the work of Nancy Goldberger et al. (1996) which researches women's views of their

own states of knowing, against the artificially simple exercises of mainstream epistemology.

This brings in its wake an emphasis on the cultural aspects of knowledge and epistemological communities as the agents of knowing (Nelson 1993). Alcoff (1996) sees this as part of a paradigm shift, now gathering momentum, towards a more socially informed epistemology. Hence I am arguing that feminist styles of philosophy are much more culturally grounded than most of their mainstream equivalents and so for this reason can be used alongside more sociologically empirical research to arrive at similar conclusions on the nature of their subject matter. For these reasons, I hope that my account of the way in which gender is inscribed in AI, whilst not being the only story that could be told, is both an interesting and plausible story.

6

FEMINIST AI PROJECTS AND CYBERFUTURES

Feminist research can have a pessimistic cast. In charting and uncovering constructions of gender, it invariably displays the way in which the masculine is construed as the norm and the feminine as lesser, the other and absent. This work is no different in that respect and I am aware of the downbeat note on which my previous chapter ends. But as both Tong (1994) and Wajcman (1991) argue, feminism is a political project and the best research is where action proceeds from description. Taking that on board for the present project involves not just using feminist approaches to criticize, but also the more difficult task of thinking through the ways in which AI research *could* be informed by feminist theory, and I make some suggestions below as to the form such research might take.

A second part of that action concerns the question of locating an appropriate feminist response to the burgeoning interest in the cultures surrounding intelligent information technologies. This includes not only AI but also the currently fashionable technologies of Virtual Reality (VR) and the Internet, both involving and related to longer established techniques from AI. Here the issue is marrying the analysis of the preceding chapters to the areas of intelligent software technology which are currently exciting considerable levels of commentary. The challenge then becomes charting a course between the Scylla of a 'nothing changes' pessimism and the Charybdis of a gushingly unrealistic 'fabulous feminist future' (Squires 1996).

FEMINIST AI PROJECTS

The fact that AI projects consciously informed by feminist concepts are thin on the ground is hardly surprising (but see e.g. Metselaar 1991). Having set up a few small projects over a period of years I have found myself questioning just what I was trying to do. I knew I was not trying to somehow 'convert' male colleagues to my way of thinking. I have never seen either my own work, or the mass of feminist literature I have consulted along the way, as proselytizing attempts to convince recalcitrant

men. I can understand how feminist writers who elicit the popular response of 'that won't convince many men', are irritated by the naivety of such comments and the way they miss the point of their endeavour. But women academics working in technological departments face pressures either not to do such work at all or only to address certain aspects. These pressures can range from whispers of 'not exactly mainstream' (which because it is a whisper I mishear as 'not exactly malestream') to actually being told not to pursue such work if they want to maintain their career prospects.[1]

Almost the only kind of work which attracts a level of respectability for women working within science and technology departments, at least in the UK, involves WISE (women into science and engineering) type attempts to attract more women and girls into the subject area; for instance, I have found male peers puzzled if I do not make myself available for university-run women into science and engineering workshops. 'I thought that's what you were interested in.' This is the acceptable face of liberal feminism (Henwood 1993) where the *status quo* is left unchallenged, where women constitute the problem, for not entering computing in the numbers that they should, and where almost any attempt to boost student numbers in an underfunded and overstretched university environment is seen as a good thing.

However those of us not prepared to wear the acceptable face of feminism return to our 'not exactly malestream' projects. Those who do projects such as these are making a statement; namely that this is research that matters, that deserves to be taken seriously and that its qualities should be judged on its own merits. And this takes more courage than many of us could reasonably be expected to muster, given the pressures I describe, and the fact that many do not have the luxury of permanent 'tenured' positions in their institutions.

If such work is not undertaken in the spirit of evangelism neither does it properly fit the notion of the successor science of the standpoint theorists (Harding 1991). This is because it is not trying to build an alternative 'successor' AI. It is, rather, and more modestly, showing ways in which AI can be informed by feminist theory and can be used for feminist projects. As Jansen (1992: 11) puts it so colourfully, it is in the spirit of 'feminist semiological guerrilla warfare . . . to transform the metaphors and models of science'. Additionally, paraphrasing Audre Lorde's (1984) metaphor it would be nice 'to demolish the master's house with the master's tools.'[2] This requires a great deal of imagination. Undeniably there are contradictions. I am reminded of the occasion when a man asked at a gender and technology workshop, 'How would a fighter plane designed by a feminist look any different?'[3] If my immediate response would be that feminists do not design fighter planes then perhaps I should acknowledge that feminists do not design AI applications either. But this will not

do as it loses sight of the political project. Hoping for change means showing how change can be made no matter how modest the beginnings.

The projects I describe below are indeed quite small. Such projects do not attract research funding and must often be tackled within the confines of final year undergraduate and masters (MSc) level dissertations. This means that individual projects are short and continuity between one project and another is difficult. I also want to make it clear that my role in these projects was as originator and supervisor, and that the results and many of the ideas and novel questions which emerged belong to the individuals who tackled the projects, most notably Chloe Furnival (1993) for the law project and Maureen Scott (1996) for the linguistics project, both of which are described below.

Some interesting problems emerge. Almost all of the students who have attempted the projects are women; the one man who built some software for teaching the history of the First and Second World Wars had to remind *me* that I had originally cast the project in terms of achieving a less gender biased approach to teaching history. As the project proceeded, I had unconsciously assumed that he was not really interested in the gender aspects, and had mentally 'written them out' of his project for him – hoist by my own petard. The women who have worked on these projects are computing students, though several are conversion masters degree students who have a humanities or social science first degree, and who generally have little background in feminist theory. There is no doubt that this makes for a difficult project, for not only do I ask that they get to grips with a new subject matter, but also it is a subject matter which requires a way of thinking completely different from the technical paradigm within which they have begun to work. In addition they are often expected to apply this to the production of a software model. But it is interesting and heartening that they invariably become absorbed by the feminist literature and usually have to be persuaded not to read any more, to get on with the business of pulling the project together. Apart from anything else it allows me to relive the excitement of my own arrival at feminism.

AI AND FEMINIST LEGAL THEORY

One of the most fertile areas for research into AI applications in recent years has been the law (see e.g. Bench-Capon 1991). Part of the appeal of the law is the way that, on the surface, legal statutes appear to offer ready-made rules to put into expert systems. A 'pragmatist/purist' debate has crystallized around this issue. Purists (e.g. Leith 1986) argue that there are no clear legal rules, the meaning of a rule is made in its interpretation, and that legal rules are necessarily and incurably 'open-textured'. We cannot know, in advance, all the cases to which a rule should apply, hence its meaning is built up through its interpretation in courts of law.

A good example, which illustrates these difficulties, was reported in the British media as I was considering this question. A woman who wished to be inseminated with her dead husband's sperm had taken her case to the High Court. Before he died the couple had been trying to have a baby. They had discussed a different case where sperm had been extracted from a dying man to inseminate his wife, and agreed that they would do the same if ever in this position. Tragically the man fell ill with bacterial meningitis. His sperm was extracted by physicians as he lay dying. However a High Court ruling was made that she could not be inseminated because, crucially, her husband's signature was never obtained; it could not have been, as he was in a coma when the sperm was removed. Mary Warnock, architect of the relevant legislation, stated that the committee which drafted the Human Fertility and Embryology Bill would certainly have permitted this case, but had never foreseen that a case like this would occur and so had not allowed for it in the statute (see the *Guardian*, 18 October 1996: 1).

Pragmatists, as the name suggests, believe that it is possible to represent legal rules meaningfully, although it is hardly a trivial task. Unsurprisingly pragmatists tend to be drawn from the ranks of computer scientists who favour predicate logic and its variants for the representation of truths in the world. Either way, it can be argued that legal expert systems embody traditional views on jurisprudence, by analogy with prior arguments on traditional epistemology and expert systems.[4] Just as feminist epistemology offers a challenge to traditional epistemology, so too does feminist jurisprudence offer a significant challenge to more traditional forms of jurisprudence. The aim of the project I describe here was to build a legal expert system to advise on UK Sex Discrimination Law founded on principles from feminist jurisprudence. It was envisaged that this system could be used by individuals, many of whom would be women, who would have little knowledge of this area of the law or of past cases which might resemble their case. Was the end product informed by these principles distinguishable from an equivalent project not founded on these principles? As the scale of the project was such that the end product was never used in a practical setting, it is not possible to answer this question definitively. In any case I argue that it was the path to the product, the journey not the destination, which was important in acting as an example of an AI informed by feminism.

Although developing in parallel ways, feminist jurisprudence appears a more practically orientated discipline than much writing in feminist epistemology, in its aim to integrate legal theory with political practice. Both disciplines have moved on from exposing violations of equal rights and sexist biases to become mature philosophical disciplines in their own right. In thinking about the women's movement in relation to the law, two areas stand out. First, there is women's use of the law to promote their

rights, with the achievement of often partial liberal measures ironically reinforcing women's oppression rather than undoing it. Second, there is the potentially more radical effort of feminist jurisprudence, which seeks to question the naturalness of legal power and knowledge, foundational beliefs about the law, and the way that legal reasoning transforms the imagined examples from male lives into a form of doctrine taken to be objective and normative (MacKinnon 1982; 1983; Grbich 1991).

Furnival (1993) points out that UK Sex Discrimination Law provides a good example of the use of these ideas in practice, particularly when we note that it is up to the individual to prove that her rights have been violated (Smart 1989: 144–6; Palmer 1992: 6). Linda Sanford and Mary Donovan (1993: 200) argue that many women have so little sense of themselves as persons with rights, that they experience considerable difficulty in recognizing when their rights have been violated. Other women may recognize that their rights are being transgressed in some way, but cannot bring themselves to make a complaint as this might brand them 'troublemakers'. Under the circumstances, any computer system designed to advise women on this area of the law would have to be presented as an unthreatening adviser which could show a client that she may have a case by analogy with past cases. The balance is important. It is unfair to offer users hope of legal redress for hopeless cases as the process of making and winning a case rests on an existing order, no matter how feminist the principles on which the system was built. On the other hand, offering examples of past cases which bear some resemblance to the present case leaves the question of whether or not to proceed open to the users, rather than making a decision for them. It is important not to make too grand a claim for what is, after all, a modest piece of work and this recognizes that considerably larger resources would be required to test out the hypotheses contained in this research.

FEMINIST COMPUTATIONAL LINGUISTICS

Given the growing interest in gender and language, computational models of language provide a potentially fertile ground for feminist projects. If feminist linguistic models challenge the models of traditional views of language, then how might this challenge be incorporated into the design of an AI system which analyses language? The project reported in this section sought to add a gender dimension to software tools which model conversational analysis (Scott 1996). This involved criticizing and augmenting a model of the repair of conversational misunderstandings and non-understandings (Heeman and Hirst 1995; Hirst et al. 1994; McRoy and Hirst 1995). The end product of the project was a formal (i.e. logic-based) model which could potentially be used to predict the outcomes of

inter-gender miscommunications, and which forms the basis for a design of a computer system which could be built to perform the same task.

Why should anyone want to build computational models of language? There are a number of reasons why the ability to represent natural language in a computer system would be desirable. First of all, a highly useful application could be found in providing natural language interfaces to existing computers systems, e.g. spreadsheets, databases, operating systems or indeed anywhere where it is currently necessary to know a series of commands. Automatic abstracting, automatic translation, intelligent language based searches for information – all these hold promise.

Part of the process of understanding language is to understand when there has been a misunderstanding between speakers and to repair that misunderstanding in a meaningful way when it occurs. This is, once again, suggestive of Collins's (1990) and Suchman's (1987) assertions that the reason that machines do not share our form of life rests upon the 'interpretative asymmetry' which exists in the interactions between humans and machines. Human beings are good at making sense of the bits and pieces of ordinary conversations, the half sentences, the 'ums' and 'ers', and so on; so good that they can make sense of almost anything and they are not easily put off. As yet, computers do not have this ability and until they do, an asymmetry in the ability to interpret utterances will remain. Hence a computer system which had some ability to repair natural language misunderstandings would clearly be of benefit in tackling this asymmetry in interpretative powers. However, the point is whether or not it is realistic to believe that a machine that can understand natural language is possible. Clearly some, such as Searle (1987), Dreyfus (1992) and Collins (1990), do not regard it as realistic. But even if, by analogy with their arguments, a full natural language understanding system might not be a possibility, then, just as expert systems can still be useful where we provide much of the nexus of understanding and background knowledge, so too could a partial natural language-processing interface be of considerable interest.

The project was inspired by an example of the finessing away of 'social factors' which is such a pervasive feature of AI and computing in general. In putting together their model of conversational misunderstanding, Graeme Hirst and his colleagues (Hirst et al. 1994) appear to have removed the subtle nuances which made the interaction into a misunderstanding in the first place. The aspect which I examine here relates to gender. Yet there are clearly many others. Race and class are two obvious dimensions; age and size are two others. This is another situation in which embodiment is important, because, of course, the speakers are bodied individuals interacting in all sorts of physical ways connected to their linguistic utterances. For instance, the following reported misunderstanding (ibid.: 227) involves, at the least, age and gender.

Speaker A: Where's CSC104 taught?
Speaker B: Sidney Smith Building, room 2118, but the class is full.
Speaker A: No, I teach it.

Hirst (ibid.) describes how the misunderstanding occurs. Speaker B assumes that A's plan is to take course CSC104, when in fact her plan is to teach it. However a number of salient facts within this example are not revealed by reading the written text alone. At the time of the reported misunderstanding, speaker A was a graduate student, and in her twenties, while B was a male administrator. Age seems to have had something to do with the misunderstanding: speaker A was young enough, and female enough, to be mistaken for a student.

An older speaker A might or might not have had the same problem – perhaps she would have been mistaken for a student's mother instead! It is interesting to speculate, in a society which values signs of youth in women, whether there might be some value in attempting to gain authority by appearing older. But this only serves to show how complex is the relationship between gender and age. True, A as a young man might have the same problem. But I wonder if a middle-aged male A would have fared differently. And what about the gender of B? The mantle of authority which men assume as they grow older is much harder for women to acquire. Women may be perceived as 'menopausal', which in Western society is almost always seen as pejorative rather than authoritative in middle life.[5] There are different ways of not taking a woman seriously which may vary according to her perceived stage in life. Hence I argue that the meaning of the misunderstanding is not readily available to us unless we have some means of reading between the lines in this way.

The large body of literature on gender and language which now exists provided a useful backdrop against which to locate this project. Chapter four noted that Spender's (1980) and Lakoff's (1975) work exerted considerable influence in the assertiveness industry of the 1980s. However, for this example, a much more pertinent body of work can be found in Deborah Tannen's research (1988; 1992; 1994), some of which is aimed at a more popular market. Most pertinently, *You Just Don't Understand* (Tannen 1992), demonstrates the sheer complexity of male and female linguistic interaction. Coupled with this, Pamela Fishman (1983) suggests that there are a number of interesting features about the way that men and women approach a conversation. She argues that women put in much more effort than men in initiating and maintaining a conversation. She also maintains that women are most often the agents of repair in misunderstandings in mixed (i.e. between men and women) conversations. If this is the case, then there is a good argument for a natural language understanding system which aims to repair speech understanding, to look at women's models of repair, if indeed they are the experts.

The complexities of men's and women's linguistic interactions are such that it seems impossible to uncover the layers of meaning in conversational misunderstandings in a model which is gender blind. For instance, Tannen (1992) offers a number of examples of misunderstandings which can only be made understandable in the light of the genders of the participants.

Hirst and his colleagues' research on the analysis of mis- and non-understandings includes a number of top-level action schemas which are used to describe the actions of the parties in a conversation. These include things like accept-plan(NameOfPlan) which signals the speaker's acceptance of the plan and reject-plan(NameOfPlan) which signals that the speaker is rejecting the plan which is being offered by the other speaker. These top-level schemas are decomposable into surface linguistic actions.

Combining Tannen's (1992) analyses with Hirst's research (Hirst et al. 1994), Scott (1996) suggests that there are a number of distinct patterns in the forms of female to female, male to male and mixed conversations so that a predictive model can be developed, that is, she claims that it is possible to predict the response expected to each form, following particular gender patterns. As women work harder to maintain a conversation, this suggests that a woman will avoid terminating a conversation using reject-plan as a man might do; instead she might use postpone, expand or replace to elicit another response from her conversant. With this revised format, Scott was able to produce more exact analyses of a number of conversations. Using the new model in the design of a conversation analysis tool gives a potential for misunderstandings to be predicted. Knowing the genders of the conversants, if a man responds with a form that is not expected by a woman, or *vice versa*, an analysis tool would recognize the beginnings of a misunderstanding possibly even before the participants can.

In this description, I am aware of the dangers inherent in suggesting that women's and men's linguistic interactions follow universal patterns. This is clearly not the case. Indeed the model described here is a white, middle-class, Anglo-American English one, which probably does not even fit, for example, New York Jewish speech, where interruptions are more common (Tannen 1994). It cannot be claimed that the model would suit cultures outside those for which it was designed. Yet making the cultural roots of the model explicit serves to underline the difficulties of generalizing linguistic misunderstanding.[6]

CONTRADICTIONS AND POSSIBILITIES

In reporting these two projects I am aware of unresolved contradictions. The computer systems that were designed and built were just as disembodied and unsituated, relying on the same symbolic representation structures as those I have criticized in preceding chapters. In going through