Social and Epistemological Bases of Technology Transfer: The Case of Artificial Intelligence

A thesis submitted for the degree of Doctor of Philosophy

by

Janet Heather Vaux

Centre for Research into Innovation, Culture and Technology,
Brunel University

December 1999

ABSTRACT

Social and Epistemological Bases of Technology Transfer: the Case of Artificial Intelligence

This thesis addresses a problem in the literature on technology transfer of understanding the local appropriation of knowledge. Based on interpretive and analytic traditions developed in Science and Technology Studies (STS) and ethnomethodology, I conceptualise technology transfer as involving communication between discursive communities. I develop the idea of 'performance of community' to argue that explanations of research and technology, and readings of those explanations, are sites for the elaboration of the identity of a discursive community. I explore this approach through a case study in the field of artificial intelligence (AI). I focus on what I call 'explanatory practices', that is practices of describing, identifying and explaining AI, and trace the differences in these practices, according to location, context and audience. The novelty of my thesis is to show the pervasiveness of performance of community within these explanatory practices, through showing the differences in the claimed identity and significance of AI, associated with different locations, contexts and audiences.

I draw out some of the implications of my approach by counterposing it to a theory of technology transfer as the passing of neutral units of information, which I argue is implicit in a complaint made by AI vendors that the AI marketplace had been damaged by overselling or hype. In particular, I show that disclaimers of hype (more than the perpetration of it) had always been associated with the marketing of AI. More generally, my claim is that it is politically important to understand that neutral information is not available even as an ultimate standard, and that the local appropriation of knowledge is not an aberration to be controlled, but a component of both successful and unsuccessful communication between discursive communities.

CONTENTS

Chapter One: The interdisciplinary context of discussions of technology transfer.

1.0 Introduction Page 6

1.1 Evolutionary economics, uncertainty and the evaluation of technology Page 11

1.2 Dissemination, local knowledge and social groups Page 16

1.3 Conclusion. Technology transfer as communication between discursive communities Page 21

Chapter Two: The case of artificial intelligence, methodology and material

2.0 Introduction Page 24

2.1 Methodological issues of a discursively based analysis Page 25

2.2 Case study: the exploitation of AI Page 29

2.3 Ethnographic standpoint and the availability of material Page 32

2.4 Conclusion. Explanatory practices Page 37

Chapter Three. How the AI industry construed 'the Problem of AI'

3.0 Introduction Page 44

3.1 The AI market: identifying growth and crisis Page 47

3.2 How AI vendors negotiated the problem of AI Page 53

3.3 Explaining the problem of AI Page 61

3.4 Misdescribing and misidentifying: toward a theory of technology transfer? Page 66

3.5 Summary Page 67

Chapter Four. Narrative and audience: AI acquires a History

4.0 Introduction Page 69

4.1 The first AI conference, the first AI program: history for the general reader Page 72

4.2 'The first AI program' and the scepticism of insiders Page 79

4.3 Conclusion. Audience as a socio-rhetorical mechanism Page 85

Chapter Five. United around the symbolic? Alliances in the field

5.0 Introduction Page 88

5.1 AI as a cognitive science Page 90

5.2 From 'AI' to 'symbolic AI': history and power Page 96

5.3 Conclusion. Page 103

Chapter Six. Reading and interpretation: ascribing intelligence
6.0 Introduction Page 107
6.1 Reading and interests: the AI debate Page 108
6.2 Interpretive flexibility: 'intelligence' as an explanatory resource Page 116
6.3 Conclusion Page 124

Chapter Seven: Industrial and political strategies
7.0 Introduction Page 128
7.1 What computers can do: non-numerical computing for an industrial audience Page 130
7.2 The Fifth Generation as the future of computing Page 137
7.3 The importance of not being extreme Page 145
7.4 Conclusion Page 153

Chapter Eight: Analytic and theoretical conclusions
8.0 Introduction Page 154
8.1 The analytic and methodological framework Page 154
8.2 Major conclusions of the case study Page 156
8.3 Contributions to the discussion of technology transfer Page 160

Bibliography Page 165

Acknowledgements

With many thanks for their comments and criticism to my supervisors, Steve Woolgar and Michael Lynch. I am grateful to Geoff Cooper for reading and commenting on Chapters One and Two. Thanks also to participants of CRICT lunch meetings who provided critical feedback on early versions of several chapters of the case study. The material and some of the arguments in this thesis were also presented in papers at a University of Bath workshop (*Humans, Animals, Machines)* in July 1995, and a University of Surrey workshop (*The Transformation of Knowledge)*in January 1999; thanks to participants at both those workshops for criticism and comments.

CHAPTER ONE
THE INTERDISCIPLINARY CONTEXT OF DISCUSSIONS OF TECHNOLOGY TRANSFER

1.0 Introduction

I begin, in this section, by exploring some of the main issues of technology transfer as a preliminary to a selective and critical literature survey in sections 1.1 and 1.2. In referring to technology transfer in this thesis, I mean a set of questions to do with the exploitation of science and technology and the promotion of links between industry, academia and government.[1] Technology transfer, in this sense, has been an increasingly important matter of political concern (realised mainly through research funding policies), at least since the 1960s (Faulkner and Senker,1995; Oakley and Owen, 1989; Tisdell, 1981; Webster and Etzkowitz, 1991). At the same time, there is some agreement that the conditions of production of science and technology are changing and reflect a closer industrial interest in research, with knowledge increasingly produced outside traditional university department structures (Etzkowitz, Webster and Healey, 1997; Faulkner and Senker; Gibbons et al, 1994). Discussions of technology transfer have developed within a number of disciplines, including economics, management studies and policy studies. There are relatively few sociological studies of technology transfer, but arguments in science and technologies studies (STS) about the relationship between the social and the technical are relevant to the discussion. In this section I explore the issues as they have been identified mainly in non-sociological writings, and conclude by noting some discussions about the possible relevance of STS ideas to these issues.

The issues addressed in academic studies of technology transfer extend over a range of topics and disciplinary perspectives, and are frequently located in discussions of technological innovation. They include (among other discussions) two major strands or sets of questions to do with causality and modelling: whether and/or how technological innovations can be said to have economic or social consequences; and how the transfer of technology may be modelled. For example, many economics-

[1] This is largely an issue of 'developed' economies. It contrasts, in particular, to studies of technology transfer between 'first' and 'third' world economies, which focus on issues such as the development gap (cf Yearley, 1988).

based studies (particularly among 'alternative' or 'evolutionary' economists)[2] start from the argument that technological advance causes economic change (Nelson and Winter, 1977, 1982; Freeman, 1984, 1988a and b; Perez, 1983; Freeman and Perez, 1988; Dosi, 1982, 1988). Approaches to modelling technology transfer, often from a policy studies perspective, include knowledge-flow models (eg Faulkner and Senker) and triple helix models (eg Etzkowitz, Webster and Healey). Questions of causality and of modelling are interlinked and counterposed in a number of ways. For example, many authors begin with a rejection of the so-called 'linear model' (where technological products are conceived as flowing in one direction from research to implementation) and discussions of the 'impact' or 'effect' of technological innovation may be criticised as involving at least an implicit linear model (Faulkner and Senker; Newby, 1992). On the other hand, the perhaps rather obvious assumption that technological innovation is economically significant underlies almost all approaches to the topic, but is only explicitly argued in some (mostly economic) texts (eg Nelson and Winter, and other authors mentioned above).

The idea that technological innovation brings economic change has been developed to suggest the idea of 'strategic technologies' that are particularly critical for economic growth. Strategic technologies are sometimes theorised in terms of Kondratiev 'long waves'[3] (Freeman, 1982, 1984) or 'technological paradigms'[4] (Dosi, 1984; Freeman and Perez, 1988; Perez, 1983, 1985). However, the idea of strategic technologies is

---

[2] This approach to economics is broadly neo-Schumpeterian (cf section 1.1, below). Most recent discussions usually describe the approach as 'evolutionary economics', a name which (as far as I am aware) was coined in Nelson and Winter's (1982) book, *An Evolutionary Theory of Economic Change*. However, other traditions have fed into the modern literature, including approaches based in the idea of Kondratiev 'long waves' (eg Freeman, 1982; 1984) which are sometimes described as 'alternative' or 'new' economics.

[3] Long waves are economic cycles, and it is a matter of dispute among economists whether these cycles are detectable in economic data (cf Freeman, 1982; Miles and Robins, 1992) . Although Kondratiev's name is always associated with long wave theory, it is Schumpeter's (1939) use of the idea which is influential. (cf Freeman, 1982, pp 207-8, who also suggests that Kondratiev didn't even invent the initial idea which should be attributed to a Dutch Marxist called van Gelderen). Miles and Robins (1992, p 8) suggest that the Freeman/Perez version of Kondratiev long waves is the most robust of recent conceptualisations: 'Long waves are now interpreted, if considered at all, in terms of the succession of "new technology systems" or techno-economic paradigms' associated with 'technological revolutions".'

[4] The idea of technological paradigms was introduced by Giovanni Dosi, inspired by Kuhn's (1962) idea of scientific paradigms. Dosi (1984, p 14) says: 'In broad analogy with the Kuhnian definition of a "scientific paradigm", we shall define a "technological paradigm" as a "model" and a "pattern" of solution of selected technological problems, based on selected principles derived from natural sciences and on selected material technologies.' The idea was further developed by Carlotta Perez in the idea of a 'techno-economic paradigm', focusing on the ability of socio-economic organisations to match the style of the paradigm (Perez, 1983; 1985). Freeman and Perez (1988) relate techno-economic paradigms with a version of Kondratiev's long wave theory.

politically effective beyond its value for economic theory, as the following passage (the opening sentence of Martin and Irvine's influential book Research Foresight (1989, p 1)) indicates:

> There is now *remarkable international consensus* that emerging technologies within the fields of electronics, information and communications, advanced materials and biotechnology will have a revolutionary impact on both economic activity and society more generally over the coming years. These generic technologies are *in the view of many* (eg Freeman and Perez, 1988; Hirooka, 1986) driving a new Kondratiev 'long wave' of economic development. [Emphases added]

One of the problems of strategic technologies is the problem of prediction, or of identifying which technologies will turn out to be strategic. Martin and Irvine, and others who have taken up the idea of foresight (particularly in the UK government's Foresight initiative), attempt to dissolve the problem of prediction through the notion of the 'foresight process' (cf Henkel et al, 1999, p 190), that is, discussion of alternative future scenarios as the context for identifying a range of possible strategic technologies (rather than attempting to predict the technological future in detail). This is a more pragmatic approach than informed, say, the fifth generation computing programmes[5] of the early 1980s which identified a single paradigm, artificial intelligence, as the future of information technology. Nonetheless, although it may provide a means of dealing with the difficulties, strategic scenario planning does not entirely dissolve the problem of prediction, and it leaves untouched the question of how technologies are evaluated (how actors judge technologies).

Other approaches to technology transfer do not particularly prioritise strategic knowledge, but ask instead about the relations between different sites of knowledge, in respect of any sort of exploitable knowledge. For example, in a triple helix model, the interest is in analysing new forms of knowledge production as a co-production of government, industry and academia. On this approach it is argued (eg Leydesdorff and Etzkowitz, 1996, p 21) that new collaborative institutional forms, which are removing the old boundaries between industry and university, act to dissolve the problem of technology transfer:

> Knowledge is no longer transferred, but co-developed.

However while it may indeed be true that some specific local boundaries are weakened or even disappear, and while from a normative point of view this may be

[5] I discuss fifth generation computing programmes at several points, but especially in Chapter Seven.

the best way to do it, it is less clear that it dissolves the general problem. Even if the distinction between universities and industrial research labs were to disappear, there will still be different social groups with different interests in the knowledge under production.[6] The question would then still arise, how do these different groups relate or communicate? Furthermore, the problem of how different groups relate is connected with the question of how technologies are evaluated. This is evident in the way in which Leydesdorff and Etzkowitz attempt to dissolve the problem of technology transfer. The co-development of knowledge, they suggest, means that 'code' is shared. And they explicate the idea of 'code sharing' by contrast to a situation where code is not shared, in an 'older model' of university-industry relations (p 21):

> [The older model] assumes that each of the partners will assess the collaboration and negotiation in terms of its own code.[7]

By contrast, newer models of industry-university relations are said to be dynamic, unstable and reflexive (p 23):

> The future location of research is expected to be found in the interaction among different contexts.
>
> The triple helix is thus explicitly presented as a destroyer of group identity.

Technology is no longer 'transferred', because it is produced in temporary collaborations which constitute both researcher and client. This conceptualisation partly depends on transforming any ideas of local interests, values and knowledge (and so on) into local 'codes', a trope that implies that social change is tantamount to rewriting computer code. It may be asked whether this does not too easily dismiss the differences between social groups in the context of technology transfer. By contrast, a knowledge flow model (eg Faulkner and Senker) retains the notion of different sites of knowledge in a way that implicitly retains the importance of social groups, and also recognises a problem of communication between groups. However, because this problem is theorised in terms of a typology of knowledge, it does not provide an obvious way of problematising local interests, values and knowledge (I return to this point and argue it in more detail in section 1.2).

---

[6] Leydesdorff and Etzkowitz themselves emphasise (p 22) that 'a triple helix is by nature unstable'.

[7] They spell out the relationships of the older model (pp 21-22): 'For example, a university department has to balance its relations with relevant partners against teaching obligations, high-level publications, and other academic objectives. The industrial partner is interested in the transfer of insights in terms of strategic and operational profits from the perspective of the business, while government is expected to orchestrate, but not to intervene in this collaboration. Thus, each partner assesses the collaborative efforts in terms of its own institutional codification.'

Some of the problems that I have identified in this section, including questions of the evaluation of technology, of communication between social groups and of social change, are reflected in sociological discussions in science and technology studies (or STS). In particular, problems of technological evaluation may be approached through discussions about the social construction of technological identities which already raise questions of relations between social groups (cf the discussion in section 1.2, below). The affinities between STS and some other approaches to technology transfer have been noted in the literature. Faulkner and Senker (p 26) suggest that:

> Policy-oriented research on industry-PSR linkage grew out of the sister fields which emerged during the 1960s - science studies and science policy.

Van den Belt and Rip (1987, p 136), writing from an STS perspective, claim:

> [B]oth the Nelson-Winter-Dosi model and the social constructivist approach developed by Pinch and Bijker share a common inspiration, which can loosely be called evolutionary thinking.

There have been several attempts to show links between evolutionary economics and actor network theory (ANT).[8] Books such as *Technological Change and Company Strategies* (Coombs et al, 1992) bring together STS authors (such as Donald MacKenzie and Michel Callon) with researchers from a variety of policy, management and economics backgrounds.[9] The editors present the alliance in terms of congruences particularly between evolutionary economics and actor networks (p 8):

> [W]here the evolutionary economist sees a stable natural trajectory, the sociologist sees a normalised irreversible techno-economic network: where the evolutionary economist sees a radical innovation, the sociologist sees ruptured networks and the emergence of new networks.

However, the claimed alliance is not limited, on the STS side, simply to actor networks. In a paper in the same volume, Callon (1992, p 72) talks of the need for showing mechanisms of both 'irreversibilisation' and 'reversibilisation', citing as an insight into the latter, social shaping and social constructionist approaches in STS (specifically Bikjer et al, 1987; MacKenzie and Wajcmann, 1985).

---

[8] Actor network theory is an approach within STS associated with Bruno Latour, Michel Callon, John Law and others.

[9] Callon has also contributed to more specifically economic collections; for example he has a paper on "Variety and irreversibility in networks technique conception and adoption" in a volume called Technology and the Wealth of Nations (Foray and Freeman, 1993).

In the following two sections of this chapter I survey some discussions selected from the literature as being pertinent to the question of a sociological (specifically an STS) contribution to discussions of technology transfer, and which also may illuminate some issues already noted in this section, to do with the evaluation of technology and communication between social groups. In section 1.1, 'Evolutionary economics, uncertainty and the evaluation of technology', I explore the way in which the uncertainty of technological development is located within the context of an economic analysis of the impact of technological innovation. I ask how this compares to accounts (in the STS literature) of the uncertainty of technological development based in the social shaping of technological identities. In section 1.2, 'Dissemination, local knowledge and social groups', I ask how the same sort of social shaping arguments bear on problems adduced in policy study models of technology transfer. I then critically question the social shaping analysis in terms of a challenge to the priority of social-group identity, made by Latour (1987). I conclude, in section 1.3, with a preliminary exploration of the idea that technology transfer may be conceptualised as communication between discursive communities.

1.1 Evolutionary economics, uncertainty and the evaluation of technology

In the previous section I suggested that contemporary interest in how technological innovations are diffused is based on a prior assumption about the wealth-creating power of new and advanced technology. This assumption is now ingrained in UK and other national policy documents, but appears to be a relatively new idea both in policy-making and in mainstream economics (cf Rosenberg 1994).[10] The power of technological advance to shape economies has been explicitly argued by alternative and evolutionary economists (Nelson and Winter 1977, 1982; Freeman 1984; 1988a and b; Perez 1983; Freeman and Perez 1988; Dosi 1982; 1988; Coombs et al, 1987; 1992) The identification of technological advance as a cause of change may be open to the criticism, made for example from a 'social shaping' perspective (Bijker et al, 1989; MacKenzie and Wajcman, 1985), that it fails to grasp the importance of the

[10] The economic historian Nathan Rosenberg comments at the beginning of his book, *Exploring the Black Box* (1994, p 9): 'It is no longer necessary for an economist to apologise when introducing the subject of technological change. That is, in itself, a (modest) cause for celebration, since the situation was very different as recently as forty years ago. ... Although sympathetic readers of Marx and Schumpeter had learned to attach great importance to technological change as a major impulse - perhaps the major impulse - in generating long-term economic growth, such an awareness had not yet rubbed off on the dominant academic traditions of western economics.'

social assignation of meaning in determining whether and how technologies are taken up. In this section I begin by looking at the way that issues of assessment and acquisition of technology are treated within a broad argument for the economic power of technological advance, and then explore some issues of uncertainty raised by a social shaping critique.

Arguments for the economic power of technological advance tend to follow Schumpeter (1942) who spoke of technological advance as introducing 'gales of creative destruction' into an economic system, providing the explanation both of destabilisation and of growth. The argument is based on a critique of 'orthodox' or 'neoclassical' economics which supposes an economic equilibrium. Neo-Schumpeterians argue that economic systems are prone to discontinuities and they identify technological advance as the cause. In this brief outline of the argument, I shall rely mainly on Nelson and Winter's influential book (1982), *An Evolutionary Theory of Economic Change.* Nelson and Winter attack, as the two 'pillars' of orthodox economics, the assumption that firms always act to maximise profits, and the hypothesis of an equilibrium in economic systems. Following Schumpeter, they see technological advance as the key to discontinuity; and they question traditional accounts of the acquisition of technological knowledge, with the help of concepts from Herb Simon (1947), Michael Polanyi (1967) and others. Nelson and Winter (pp 59-65) are particularly critical of the way in which orthodox economics tends to treat 'knowledge' as just one more factor of production. In the abstractions of orthodox economics, they argue, rational companies can tell what will be the outcome of their investment: companies are assumed to be 'all-knowing'. This does not mean that actual companies are supposed to know the effects of their choices, but that the market behaves as if they do (this is 'the invisible hand' at work). Nelson and Winter (pp 202-4) reply that real firms are not all-knowing; nor does the economy behave as if they were.

Following Simon (1947), Nelson and Winter replace the orthodox description of corporate behaviour (rational selection across a known set of choices) with the idea that companies search for solutions in an ill-defined search space. They emphasise (pp 171-2) that such searching will be inherently uncertain, that it will be contingent on the local historical context, and that it will be irreversible insofar as the cost of

search will mean companies have an investment in any chosen path. Here 'uncertainty' is adduced in comparison with the standard of full knowledge set by orthodox theory, but there is also a reflexive sense of uncertainty since individual firms are second-guessing the behaviour of firms in general.[11] Nelson and Winter's model of the behaviour of firms is based on the following of routines, which are good enough,[12] and which tend to be followed for as long as they work.[13] The behaviour of firms faced with the need to assess and acquire new technology is therefore crucial to economic discontinuity. In a later paper, Nelson (1987) compared technological advance to a horse race: some technologies and companies will win and some will lose; advance happens because the horse race takes place.[14]

There have been a number of responses to the ideas of evolutionary economics by sociologists working in science and technology studies (STS). Some of these[15] (Pinch and Bijker 1987; van den Belt and Rip, 1987; MacKenzie, 1990; 1992) are based on a critique of Nelson and Winter's notion of 'technological trajectory'[16] as implying technological determinism. Technological trajectories, according to Nelson and Winter (pp 258-9), are based in 'technicians' beliefs about what is feasible or at least worth attempting'. That is to say, the trajectories are described not in terms of the logic of technologies, but in terms of the behaviour of technologists. Similarly,

[11] Because the environment includes the decisions made by other firms, there is a relevant paradox to be found in game theory. As Kenneth Arrow (economist and Nobel laureate) put it in a recent book review relating to the field of game theory (*Times Higher Education Supplement*, 23 October 1998, p 23): 'Each entrepreneur or other decision-maker would base his or her current investments on forecasts of the choices made by others: but these actions were in turn dependent on the forecasts made about the original individual's choices.' For this reason, Arrow suggests, 'the Austrian economist Oskar Morgenstern and the US sociologist Robert K Merton had expressed the view in the 1930s that economic and social forecasting was impossible.'

[12] This is based on Simon's notion of 'satisficing', which he explained (1969, p 36) as a pragmatic reaction to the complexity of the economic environment: 'Normative economics has shown that exact solutions to the larger optimisation problems of the real world are simply not within reach or sight. In the face of this complexity the real-world business firm turns to procedures that find good enough answers to questions whose best answers are unknowable.'

[13] The model is 'evolutionary' because those firms that hit on winning routines are the ones that survive.

[14] From the point of view of any individual firm, however, the greater economic good is not much compensation for falling at the first fence.

[15] It has also been argued (Callon 1992, 1993; Coombs et al 1992; Foray and Freeman 1993) that actor network theory (ANT) has affinities with evolutionary economics, including a common concern with heterogeneous networks and with uncertainty.

[16] The notion of technological trajectory was introduced by Nelson and Winter to help account for the way in which technological innovation is cumulative (relatively certain) and irregular (Nelson and Winter, 1982, pp 255-262). They suggest, for example (p 262), that given the introduction of new knowledge which, say, made old skills obsolete, and introduced new ones, '[...] technical progress would surge forward as solutions appeared to problems suddenly made relatively easy by the strengthening of the knowledge base - only to slacken again as the new areas of search become, in their turn, relatively well explored.'

Giovanni Dosi, who suggests that technological trajectories are the technological equivalent of 'normal science' (cf Kuhn, 1962) following from the institution of new 'technological paradigms', explains the trajectories (1984, p 15) in terms of a blinding of the technological imagination to other possibilities . It is not entirely clear why this should be taken as an example of technological determinism, except perhaps that the technologists are seen to fail against an implicit standard of the potential of technology. However, science studies has long been alert to hints of technological determinism as part of its inheritance from the Sociology of Knowledge, particularly Bloor's (1976) Strong Programme. Bloor's approach was strongly causalistic, initiating a search for social (or other non-scientific) causes of scientific achievements. The Strong Programme challenged the idea that scientific knowledge was the outcome simply of a search for truth. The analogous task, in the case of technology, is to challenge the idea that technological advance occurs simply from following the logic of technology. Technology is not autonomous, the argument goes, but shaped by society (cf Mackenzie and Wajcman, 1985). This argument often takes the form of the rejection of the idea that technology 'impacts' upon society.[17] So, for example, MacKenzie and Wajcman (p 8) reject discussions of the 'effects' or 'impact' of technology on society, on the grounds that technology is not independent of society in the first place:

> Our focus - and where our criticism of technological determinism would centre - is on the assumption that technological change is autonomous, 'outside' of the society in which it takes place. Our question is, what shapes the technology in the first place, before it has 'effects'? Particularly, what role does society play in shaping technology?

This emphasis on the primacy of the social over the technical is not an obvious issue for the evolutionary economists.[18] The social shaping argument, however, is not just about reversing the causal arrow.

---

[17] It is not clear, however, to what extent this idea is actually held by academic sociologists and historians of science. MacKenzie and Wajcman (1985) name Lynn White's history of the stirrup as a prime example of the offence, but they do this through a very uncharitable reading which takes White to be saying that wherever there are stirrups there is feudalism. A better example, not cited by MacKenzie and Wajcman, might be in the tradition of idealist history of ideas, for example the work of Harold Innis (1951) who argued that every culture gets a primary 'bias' from the technology it produces.

[18] In more recent writings this situation has partly changed, at least programmatically. For example, Nelson and Soete (1988, p 633) remark: 'Indeed it will be the broad societal context: including economic, but also social and ethical factors will set the conditions within which technological change will be adapted, even selected.'

More importantly, perhaps, the exploration of how the social shapes technology has led to a description of 'the social' not just at the level of causes, but also as the locus of meaning. (Bijker et al, 1987; MacKenzie and Wajcman, 1985) This means that claims about the uncertainty of technological development, in the STS tradition, imply openness, multiple meanings, and interpretive flexibility in determining what a technological product is (and what it is for). As Bijker et al (1987, p 12) put it, in the introduction to The Social Construction of Technology:

> Because social groups define the problems of technological development, there is flexibility in the way things are designed, not one best way.

Pinch and Bijker (1987) exemplify this argument with a history of the bicycle aimed at convincing the reader that there is nothing intrinsically 'better' about a modern bicycle compared to, say, a penny-farthing, nor is there any inevitability about the route from penny-farthing to modern bicycle. The location of optimal design, of what the bicycle 'means' is, they argue, to be found in various social groups which have different interests in what a bicycle is for: optimal design and technological advance is an outcome of social negotiation.

MacKenzie (1992) notes that, despite elements of technological determinism, evolutionary economics stresses the uncertainty of technological outcomes. The discussion of uncertainty in social shaping texts provides a contrast to the uncertainty described by evolutionary economists: it is a contrast between the freedom of decisive choice and the luck of the horse race, between the social negotiation of outcomes and strategies to manage the overwhelming power of circumstance. Moreover, in the radical versions of uncertainty which each discipline offers we may see quite divergent theorising: on the one hand the impossibility of rational choice, on the other openness of meanings.[19] Nonetheless, evolutionary economics and social shaping both agree on the uncertainty of technological outcomes. Evolutionary economics, however, asserts the economic significance of technological innovation at the same time as it points to the difficulty of prediction and the problems that this creates for technology strategy both for firms and for governments. Nelson and others (Nelson 1959, Arrow 1962) have argued that

---

[19] To foreshadow discussion later in the thesis (cf especially Chapter Six), this may be seen as an example of the performance of academic discipline based in the negotiability of theoretical concepts.

market forces provide inadequate resources for R&D,[20] suggesting a need for government action, at least in funding basic science,[21] if not in promoting technology transfer. There is a difference of opinion among alternative and evolutionary economists on the 'sharpness' of change brought by technological innovation, but most agree that there are important policy implications (cf Nelson and Soete, pp 634-5).[22] This provides the context for a more detailed discussion of the social, institutional and political mechanisms through which technology transfer is, or may be, enabled. The latter discussions point again to the need for a way of theorising relations between social groups, indicated in section 1.0. In the next section I look at some of the ways technology transfer has been problematised, particularly in technology policy literature, and ask how insights into the social constitution of meaning (discussed in this section) can contribute to this discussion.

1.2 Dissemination, local knowledge and social groups

I suggested previously (section 1.0), that there is a question whether models of technology transfer adequately represent social relations and communication between social groups. Models of technology transfer are often approached through a critique of the so-called 'linear model' (cf Clark and Staunton, 1989; Newby 1992; Schmoch et al, 1993; Faulkner and Senker, 1995). The linear model may be seen as reflecting the economic representation of technology as the creator of wealth (as in the texts discussed in 1.1). Indeed, the linear model is sometimes blamed specifically on Schumpeter (Newby, 1992). Arguments against the linear model generally aim to

---

[20] Cf, for example Freeman (1993, p 25): 'Ever since the seminal papers of Nelson (1959) and Arrow (1962) there has been a widespread consensus that this [the financing of fundamental research] was a clear case of market failure.' Nelson and Soete (1988, p 631) stress Arrow's role: 'Since Arrow's seminal contribution in this area some twenty-five years ago, there is general agreement that market failure is indeed one of the intrinsic characteristics in this area and that underinvestment in R&D will be the logical outcome of market allocation.'

[21] Nelson and Winter (p 391) point out that what counts as basic and applied research is socially determined: 'Society in effect has a choice regarding what arenas of research it will define as basic research to be funded publicly and guided by the tenets of a scientific discipline, and what areas it will regard as applied and to be guided (if not necessarily funded) by criteria close to the values of the organisations using particular technologies.'

[22] Nelson and Soete (1988, pp 634-635) in the concluding paragraph of a concluding chapter to a collection of papers on Technical Change and Economic Theory, remark: 'In this volume several authors, but especially Carlota Perez and Christopher Freeman, have proposed that we are entering a new era with a drastically different operative economic-technical paradigm. Not all of the authors of this book would put the matter as sharply. But all would agree that there are important new things going on and to be monitoring. Let the exploration go on. Let it be urged and supported. But new government structures and public laws will be needed to support the valuable [sic] of the new, and to constrain the pernicious.'

provide more complex channels of flow, so that information can pass two ways, feedback is enabled, and so on. In diagrammatic versions of the linear model, this can lead to increasingly baroque creations, with multiple levels and complex feedback loops (cf Schmoch et al, 1993; Woolgar, 1994b).

An alternative to producing ever more complex diagrammatic representations of dissemination is to problematise the links and mechanisms by which knowledge is disseminated, and several authors have approached this task. I have mentioned two of these approaches above (section 1.0). The 'triple helix' model describes new collaborative institutional forms and mechanisms in which knowledge is 'co-developed' (eg Etzkowitz and Leydesdorff, 1996). This (as I suggested in section 1.0) perhaps too easily removes problems of linkages and relations between social groups. The knowledge flow model developed by Faulkner and Senker incorporates the idea that knowledge has different locations and asks about the movement and transformation of knowledge between organisations.[23] In this model, knowledge has various organisational locations (such as educational institutions and corporate research labs), and the authors also provide a typology of knowledge which allows them to distinguish different sites of knowledge (such as textbooks or embodied skills) which may be more or less available (more or less explicit or tacit) and require different social forms of linkages (such as personal visits or reading of journals). This provides a nuanced account of inter-organisational linkages. However, as Sanz-Menendez and Webster (1996) have noted, this approach generates the further question of how communicated knowledge is appropriated locally.[24] Faulkner and Senker (pp 20-23), in a section on the limits to academic-industry links, mention several 'barriers', including failure to communicate, differences of agenda and 'mutual suspicion and antagonism'. Perhaps these barriers can be conceptualised through the idea that different social groups appropriate knowledge according to their own purposes or interests, following Pinch and Bijker's (1987) paper (discussed in section 1.1). Latour in his book *Science in Action* (1987), however, raises doubts

---

[23] Faulkner and Senker (1995, p 42) talk of 'knowledge', both to avoid implications of a hierarchical distinction between science and technology, and in reference to 'the overall knowledge requirements of the innovating organisation'.

[24] In a proposal for a new COST action in social science, Sanz-Menendez and Webster (1996, Section A) point to the need for a broader understanding of how firms and other organisations 'cope with and shape' knowledge flows. They suggest: 'They will share, exchange and depend on translating knowledge among themselves. But they will also do so according to their localised needs and capacities. These can be called distinct "knowledge constituencies".'

about the analytic value of 'social groups' as an explanation of problems in the uptake of science and technology. In the remainder of this section, I shall further explore these two contrasting positions on the power of social groups to constitute meanings.

Latour challenges the analytic value of the concept of social groups partly on the basis of his more general claim to have surpassed the Subject-Object distinction.[25] But it is also an element in an important attack on what he calls the 'diffusion model' of the uptake of scientific knowledge or of technological artifacts. In his critique of the diffusion model Latour rejects talk of technological impact (as do Pinch and Bijker), but he also rejects talk of social shaping, or indeed any talk of 'the social'. The diffusion model, he suggests (1987, p 141), in fact generates the distinction between the social and the technical:

> In the diffusion model society is made of groups which have interests: these groups resist, accept or ignore both facts and machines, which have their own inertia. In consequence we have science and technics on the one hand, and society on the other. ... Let us go further: belief in the existence of a society separated from technoscience is an outcome of the diffusion model.[26]

He proposes, in place of a diffusion model, a model of 'translation', which does not distinguish the social from the technical, or subject from object, but recognises only:

> heterogeneous chains of association that from time to time, create obligatory passage points.

In addition to these general arguments against the social/technical distinction, Latour also has specific arguments accounting for the construction of social groups. This may be illustrated by his tale of Eastman's 'invention' of amateur photographers. As he puts it, in a form of words designed to emphasise the novelty of the point (p 122):

> Eastman had the bright idea of inventing a new group of 6- to 96-year-olds that was endowed with a craving for taking pictures. This enlistment depended on a camera that was simple to operate ...

More formally, 'inventing new groups' is one of the Machiavellian tactics identified

---

[25] Although this perhaps is more evident in later writings, particularly *We have never been modern* (1991).

[26] Later on the same page (141), Latour spells out the implication for analyses that begin with a rejection of technological determinism: 'Social determinism courageously fights against technical determinism, whereas neither exists except in the fanciful description proposed by the diffusion model.'

by Latour in his description of the complex steps involved in the achievement of translation. This is Tactic Three of Translation Four, 'reshuffling interests and goals' (p 115).

To see the analytic force of Latour's emphasis on the invention of social groups, it is useful to compare it with an argument that makes much of the power of social groups, Pinch and Bijker's history of the bicycle. In particular, they claim (p 30):

> In deciding which problems are relevant, the social groups concerned with the artifact and the meaning those groups give to the artifact play a crucial role: a problem is defined as such only when there is a social group for which it constitutes a 'problem'.

However, the relevant social groups identified by Pinch and Bijker, as assigning different meanings to the bicycle, are themselves identifiable largely in terms of their relationship to bicycles. Pinch and Bijker continue (p 30):

> The use of the concept of a relevant social group is *quite straightforward.* The phrase is used to denote institutions and organisations (such as the military or some specific industrial company), as well as organised or unorganised groups of individuals. The key requirement is that all members of a certain social group share the same set of meanings, attached to a specific artifact. [My emphases]

The social groups actually identified by Pinch and Bijker (pp 30-42) include 'consumers' or 'users', 'anticyclists', 'women cyclists', 'young men of means and nerve' for whom 'the function of the bicycle was primarily for sport', and 'elderly men'. While one might argue that 'women' or 'young men of means and nerve' or 'elderly men' are in some sense already social groups, it is only insofar as they have a relationship to bicycles that they become 'relevant' social groups. If Pinch and Bijker suppose that bicycles only become meaningful insofar as cyclists give them various meanings, where did the cyclists (somehow already cyclists) find these pieces of metal that they then deemed to be bicycles? This is perhaps an uncharitable reading of Pinch and Bijker, whose project is to show that engineering considerations alone do not determine what counts as optimal design, that there is no 'one best way' (p 28). However, Latour makes an important point, which Pinch and Bijker at best gloss over, that technological artifacts are already socially active. They are designed to entrap allies (users) and already address (configure) potential groups

of users.[27] Perhaps it is not the case that the concept of a social group ('relevant' or otherwise) 'is quite straightforward'.

The difference between Latour on the one hand, and Pinch and Bijker on the other, is not merely one of prioritising either artifacts (engineers) or users. Each text may be seen as taking a slightly different stance on the Kuhnian project of re-opening scientific and technological facts, for the purposes of seeing how they were stabilised (Kuhn 1962, p 1). Part of the power of Pinch and Biker's paper is to do with a re-opening, or a distancing, of a familiar object. The bicycle is an excellent case to use, because it is difficult to believe that the penny-farthing wasn't doomed from the start. In convincing us that the penny-farthing (or, as it was once tellingly called, 'the ordinary') might have had a future, Pinch and Bijker reveal the whiggish hand of technological history. Latour, by contrast, aims to show us all the steps by which such a history is written, including all the 'boxes' that get closed along the way, by the various mechanisms of recruiting allies and enrolling interests. But in doing this, Latour often depends on a style of story-telling which, explicitly in his theory, is itself a mythic outcome of the networks, that is, the myth of the lone inventor (or, indeed, the myth of Machiavelli).[28] It is not clear, then, why the fact that social groups are, for ANT, an outcome of the networks should in itself make them less appropriate as analytic concepts than, say, the concept of 'power'. For everything, including power, is surely an outcome of the networks, according to ANT.

There is one more thing to notice about the social groups adduced by Pinch and Bijker, as well as by Latour, which is that they tend to be identified as social groups insofar as they are users of a technology, be it racing cycles, safety cycles, or Eastman's camera. A similar point may be made in relation to Woolgar's (1991) notion of 'configuring the user', which shares with Latour a sense of the construction of users and their needs, in this case through the technological 'text'. However, the problem of the identification and evaluation of technologies, especially strategic

---

[27] My parenthetical use of more Woolgarian terms is to indicate that this point can be made from other directions than actor network theory. For the concept of 'configuring the user', cf Woolgar (1991).

[28] The myth is generated by the invisibility of the masses of enrolled allies. Latour (1987, p 174) remarks: 'Although scientists are successful only when they follow the multitude, the multitude appears successful only when it follows this handful of scientists! This is why scientists and engineers may appear alternatively endowed with demiurgic powers - for good or bad - or devoid of any clout.'

technologies goes beyond a vendor-user relationship. Among other things it involves communication between established institutional (and quasi-institutional) groups, including academics, policy analysts, industrial strategists, funding bodies, venture capitalists, and so on. What is required is an approach to understanding social groups which goes beyond the designer/user paradigm, but does not assume that social groups are given (with their interests) from the start.

1.3 Conclusion: technology transfer as communication between discursive communities

I began this chapter by identifying two major themes in discussions of innovation and technology transfer: the first concerned with theorising the economic impact of technological innovation; the second concerned with modelling links and movements between various sites involved in the diffusion of science and technology (or, in the term now most commonly used, 'knowledge'). Each of these themes is broadly associated with bodies of literature in different disciplines: the first in economics and the second in policy studies. These themes are also somewhat indirectly reflected in sociological discussions within science and technology studies (STS) which are concerned with relations between the 'social' and the 'technical' (and I include here those, such as Latour, who argue that this is a false opposition (cf section 1.2)). On the topic of the socio-economic impact of technological innovation that interests evolutionary economists, STS texts initially seemed to offer a contrary view arguing that it is the social that impacts on the technical; but there is also a more interesting complementarity between the two literatures, which agree on the uncertainty of technological outcomes, but with divergent emphases in describing this uncertainty (cf section 1.1). The STS description of uncertainty is based in the idea of interpretive flexibility, that no one meaning or use is given with (inscribed in) a technological artifact, but that what an artifact is (and is for) is the outcome of negotiation between social groups of users, each with different interests. This seemed to promise a notion of uncertainty that might help provide a fuller description of some of the problems of technology evaluation and acquisition suggested in the economic models, and at the same time to offer a means of conceptualising local interests and 'barriers' in a way that would be useful in models of technology transfer. However, Latour (1987) points to a problem with taking

social groups as analytic 'givens', complete with interests and purposes that define what an artifact is; indeed, one may as easily claim that the artifact creates the users as a social group (cf section1.2).

What is needed is a way of theorising social groups which allows for the co-production of artifact and user, as well as explaining the persistence of local interests, and addressing the question of the local appropriation of knowledge. The notion of discursive community is useful here. I take a discursive community to be one which is identified and maintained through negotiation among members, and between members and non-members. This understanding of discursive community is based on discussions within STS which owe much to the traditions of ethnomethodology and conversation analysis (cf Lynch, 1993, for an overview). This approach provides a number of concepts for analysing discursive material (which I discuss in more detail in 2.1, below). The idea of discursive community lies behind notions such as 'configuring the user' (Woolgar, 1991), where for example help desk staff perform community ('who we are') in the formulations through which they configure their clients. In principle, at least, the rhetorical performance of community may be shown in a variety of representational practices, including the identification and evaluation of artifacts, and the construed significance of scientific research or technological skills (cf Cooper and Woolgar, 1993).[29] My working assumption, in embarking on this thesis, is that conceptualising technology transfer as communication between discursive communities (Woolgar, 1994b) provides a unified way of investigating the various questions relating to technology transfer that I have noticed in this chapter, such as how actors judge technologies, what the 'cultural barriers' are to technology acquisition, and how knowledge is appropriated locally or, in other words, the social and epistemological bases of technology transfer.

In Chapter Two, which now follows, I propose to transform the above rather programmatic statement of aims into a coherent and detailed account of the terms and intentions of my thesis. I provide a more detailed description of a discursively-based methodology (2.1), introduce my case study, on the exploitation of artificial intelligence (2.2) and describe the evidential basis (the material and the ethnographic

[29] For example, in a study of a committee defining software quality standards, Cooper and Woolgar (1993) show the co-construal of what the proper concerns are about quality and who is properly concerned with testing quality and formulations of who we are.

perspective) of my case study (2.3).

CHAPTER TWO
THE CASE OF ARTIFICIAL INTELLIGENCE, METHODOLOGY AND MATERIAL

2.0 Introduction.

At the end of the previous chapter I suggested that technology transfer might be seen in terms of communication between discursive communities. In this chapter I begin by locating this conceptualisation within the methodological traditions of STS, and ask how it relates to questions of technology transfer as raised specifically in terms of my case study. I addition I address some further methodological issues to do with sources of material and ethnographic standpoint, which are raised by the fact that some significant parts of my material were gathered, prior to embarking on a sociological study, when I was more a participant than a participant-observer in the AI industry. I conclude with a summary outline of the methodological basis of my case study, identifying the specific object of study and describing the chapter-by-chapter structure of the case study.

I begin (2.1) with a discussion of some approaches within STS to the use of discursive material and associated theorisations. I address these from a methodological viewpoint, asking what the discursive object is and what is implied in terms of going about the interpretation of discursive material. More specifically, I ask what these traditions have to offer for a study of communication between discursive communities, looking particularly at approaches which can show the negotiation or performance of community.

In 2.2 I introduce my case study, which relates to the exploitation of AI. I ask first how this case relates to the broader discussion of technology transfer; and second whether and how issues of communication and performed community are germane to this case study. I also ask whether and how substantive questions about the nature of knowledge are relevant to a study of technology transfer, by comparing some other STS approaches to AI as a theory of knowledge.

In 2.3 I ask what issues in practice or in principle are raised by my closeness to the

community which I am studying. This arises because my experience as a journalist covering the AI industry provided my initial contact with the AI community, and some significant sections of my thesis are based on material that I gathered not as a sociologist but as a specialist trade journalist. I approach these issues by looking at some anthropological discussions about the authority of the participant-observer and the difficulties of 'indigenous anthropology', and ask whether and how these debates are relevant to my own case.

Finally, in 2.4 I summarise the discussion in the preceding sections of this chapter, as a basis for exploring the appropriate discursive objects of this case study. This section concludes with a chapter-by chapter outline of the structure of my case study.

2.1 Methodological issues of a discursively based analysis

To suggest that technology transfer may be described as communication between discursive communities (as I did in 1.3) is already to locate my approach within some strands of STS that not only make use of discursive (spoken and written) material, but also derive a variety of methodological and theoretical lessons from this use. There are a number of contrasting theorisations within recent work in STS. For example, the idea of technology as text (Cooper and Woolgar, 1993; Grint and Woolgar, 1997; Woolgar, 1991) conceptualises the identification and use of technology as a 'reading', to emphasise the role of the user (the social) in determining what a technology is and what it is for. In contrast, the voice given to 'non-humans' in actor network theory is intended to rupture the social-technical divide (Callon, 1986; Latour, 1987, 1991). Elsewhere, Lynch (1993) criticises both Latour and Woolgar for distancing meaning from the local production of scientific knowledge. Such differences in theorising discursive material also inform some of the methodological choices, and the methodological and theoretical issues are not easily abstracted from each other.[30] In the following discussion, however, I attempt to emphasise the methodological aspects.

To use discursive material implies more than the mere use of spoken or written

[30] As Tsatsaroni and Cooper (1999) point out, this is particularly so insofar as STS approaches to discourse analysis are historically based within an empirical problematic of handling data, in contrast to the continental tradition.

material. The use of interview techniques is common in the social sciences but often approached in an objectivist framework, where the point is to gather data or information, and the interviewees' words are taken as a report and sometimes indeed as a report on the world.[31] More commonly, however, interview material is taken as a report of the subject's views or opinions. An example from the science studies literature is Mulkay and Gilbert's (1981) use of interviews with scientists to contrast scientists' talk of what they do with Popper's description of what scientists do. More recent approaches in STS have been more concerned with the 'performative' aspect of discursive material, although performativity is a concept which is open to varieties of interpretation (some of which I discuss below). The idea of discursive interpretation has also been extended so that the discursive object is no longer necessarily a piece of speaking or writing, but may be, for example, scientific practice or the social world. Latour and Woolgar each transforms the discursive object of analysis in a different way, Latour through a semiotically-based concept of 'translation', and Woolgar through a radical conception of the text/artifact as interpretatively open, but both implying the power of language to act (cf slightly different arguments to this effect by Lynch (1993, p 99) and by Tsatsaroni and Cooper (1999, p 14)). Do methodological implications follow from this 'linguistic turn'[32] or do the arguments attach predominantly to questions of theorisation? In the following paragraphs I ask what is on offer in practical methodological terms as an approach to discursive material, looking particularly at some aspects of the work of Latour, Lynch and Woolgar.

Actor-network theory (ANT), with which Latour has been strongly associated (in collaboration with Michel Callon, John Law and others), presents itself initially as a set of methodological guidelines. That is to say, ANT describes the achievement both of scientific knowledge and of technological advance in terms of 'translations' through which actors enrol allies. The task of the analyst is to 'follow the actors'. Guidelines as to what the novice analyst might expect along the way appear to be set

[31] For example, in the course of a research project (EPSRC Grant GR/K 32548) that involved interviews with university Industrial Liaison Officers (ILOs), I was shown a successful MBA thesis that had been partly based on interviews with university ILOs. I was struck by the fact that while we were exploring the variety of ways in which ILOs' talk configured academics, the MBA thesis used the same material as evidence of whether or not academics were aware of issues of exploitation, IPR, etc. The ILOs in the MBA thesis were treated as expert witnesses, on our (STS-based) approach, the talk of the ILOs was interrogated for the performance of social relations.

[32] This phrase was possibly coined by Rorty (1967) in his book titled *The Linguistic Turn*, to refer to the development of the philosophical movement sometimes called 'linguistic philosophy'.

out in Latour's (1987) *Science in Action*, in a detailed description of the tactics and rules of method employed by successful actors. These are followed, in the appendices, by rules of method and principles for the analyst herself to apply. However, the rules of method for actors are not actually a how-to-do-it guide for entrepreneurs or scientists, but a redescription of efficacy (scientific truth or technological advance) for the purposes of philosophical argument. Similarly the rules of method for the analyst seem to provide a philosophical rather than a methodological orientation to carrying out a case study. It is significant that in *Mapping the Dynamics of Science and Technology* (Callon et al, 1986), which does claim to offer a method for the study of science and technology, the method is co-word analysis and ANT supplies the philosophical justification.

The effectiveness which Latour describes through 'translation' is one way of approaching the performativity of discourse. In anglophone traditions there are other routes to performativity. The term 'performative' was introduced into British 'ordinary language' philosophy by Austin (1961) in the 1950s[33] who was describing a type of utterances which are not reports, but 'perform' a deed (examples include 'I promise' and 'I apologise')[34] . In the STS conception of performance, however, the ethnomethodological concept of 'indexicality' is a more important root. Garfinkel and Sacks (1970) took the name from Yehoshua Bar-Hillel[35] (1956), who used the term 'indexical expressions' for a broad class of words (such as 'here', 'now', 'she', 'this') whose meaning varied with occasion and context of use. Lynch (1993, pp 17-18) says of Garfinkel's use of 'indexicality' that it is 'simply another way of speaking of the entire picture of social order'. The question of indexicality disappears into the ethnomethodological project, in a way that leaves virtually no gap between theorisation and methodology (p 22):

> [W]hat becomes prominent is not that all expressions are indexical but that members

[33] Austin's paper 'Performative Utterances' was published posthumously in a collection of his papers edited by Urmson and Warnock (Austin, 1961). The editors, in a Foreword, say that this paper 'is a version, with minor verbal corrections, of an unscripted talk delivered in the Third Programme of the BBC in 1956'.

[34] Austin (1961, pp 233-4) argued that in abstract there is a clear distinction between reports and performances, but that in practice there is a pervasive ambiguity: 'If he had said "I feel perfectly awful about it", then we should think it must be meant to be a description of the state of his feelings. If he had said "I apologise", we should feel this was clearly a performative utterance, going through the ritual of apologising. But if he says "I am sorry" there is an unfortunate hovering between the two. The phenomenon is quite common.'

[35] Bar-Hillel was a pioneer in the field of machine translation, that is, computer translation between one natural language and another. Bar-Hillel is also important in the history of Artificial Intelligence as a recognised precursor to research in natural language understanding.

> manage to make adequate sense and adequate reference with the linguistic and other devices to hand. The question for ethnomethodology is, How do they do that?

Ethnomethodological studies also provide descriptions of how local negotiation is done that have provided a model or inspiration for some STS studies that are not directly in the ethnomethodological tradition. In particular they suggest a way of approaching the idea of the negotiated performance of community through the idea of formulations of membership (cf Sacks,1972; Schegloff, 1972). Schegloff pointed to the pervasiveness of construals of membership which can occur within conversations ostensibly on some other matter. He explored this within interchanges about (geographical) location (p 93), where the selection and recognition of formulations of location can, he claims, be seen as a basis for

> demonstrations of, claims to, failings in, decisions about, etc, the competent membership [of the 'same community'] of either speaker or hearer.

This, he goes on to suggest, provides the 'never-ending' possibility of the testing of adequate membership of relevant communities.[36] Cooper and Woolgar illustrate the performance of community through a case study of a government-initiated project to define a framework for assuring software quality. In this case, performance of community is shown by displaying the ties between decisions about what counts as 'quality' in software and negotiations of relevant community, appropriate skills, etc. The analytic point (that this is performance of community) is made through describing the rhetorical achievements of various actors. However, there is a theoretical question about the status of such descriptions. Tsatsaroni and Cooper (1999, pp 13-14) argue that the implication of performativity contained in the idea of performance of community is contrary to the main thrust of Woolgar's theorisation; and that the textual metaphor ('technology as text') prioritises the openness of the text. This seems confirmed by Grint and Woolgar (p 73) who argue that showing performance of community is only of limited value in theorising the relation between the social and the technical, since it can only reveal local productions of the social/technical divide but not go beyond it.[37] Thus while, for Woolgar, social

---

[36] Schegloff (1972, p 94) says: 'Insofar as friendships, reputations, marriages, collaborations, etc, may turn on someone's competent membership in some class of members (eg "swinger", "anthropologist", "good Jewish girl", "Africanist", etc), each occasion of the use of a place formulation selected because of its presumed recognisability to a member of such a class is part of a never-ending potential test in which persons can be shown to be inadequate members of the class, and thereby inadequate candidates for the activity.'

[37] It might be argued that this begs the question of whether there is an analytic point beyond the local production of the world. Cf Lynch's criticism (1993, pp 95-100) of Latour and Woolgar as attempting to act as observers from Mars.

relations are based in the performance of community and negotiation of boundaries, mundane discourse is transcended through a radical moment, the moment of the 'methodological horrors' (cf Woolgar, 1988a) and the 'constituting reader' (cf Tsatsaroni and Cooper, p 13): interpretation is revealed to be radically arbitrary. Methodologically this radical moment is not habitable, but suggests a meta-methodological trap, the possibility of subverting any interpretation.

Finally, in this section, I come to the question of what this suggests for a study of technology transfer as communication between discursive communities. The first implication is that the performance of community may provide the basis for a notion of discursive communities based in the negotiation of boundaries. The second is that it is methodologically appropriate to look for the co-production of community, of audience and of technological identities in the communication between discursive communities (and that this may be particularly appropriate given the institutionalisation of technology transfer mentioned in section 1.0). The details of how these possibilities may be transformed into a methodology specific to my case study are discussed in the final section of this chapter (2.4), following an introduction to my case study, and discussion of ethnographic standpoint (which raises matters of further relevance to the question of analytic 'distance').

2.2 Case study: the exploitation of Artificial Intelligence

In Chapter One I identified, among the practical themes at issue in discussions of technology transfer, the evaluation of strategic technologies and the making of links between (among others) universities and industry. In this section, I first locate my case study in terms of some of the themes at issue in practical discussions of technology transfer and then explore its relevance to a study of communication between discursive communities. Finally, I locate my thesis in contrast to other sociological studies of AI and of the proper description of knowledge.

The AI industry may be identified as making its first appearance around 1980, when it consisted mostly of small university spin-off companies (Ovum, 1984; section 3.1 of this thesis). These companies sold hardware and software products based on research that had been under way for about two decades in a small number of high ranking academic institutions, mainly in the US but also in Europe, especially the UK

and France (Fleck, 1982; Ovum, 1984). The products, initially, were mainly aimed at the development of 'expert systems'[38] and, again initially, were mostly written in an AI language such as Lisp (or, in the UK and France, Prolog). As an example of technology transfer, then, this is a case of 'knowledge-led' innovation (Senker and Faulkner, 1995, p 27), and its basis in small university spin-off companies was a form considered particularly appropriate to research-based innovation (cf Segal, Quince and Partners, 1985). US companies, in particular, were successful in gaining venture capital funds (Ovum, 1986a; 1986b). At the same time, in the early 1980s, AI was identified as a strategic technology in several national and international strategic funding programmes for 'fifth generation' computing. These included the Japanese Fifth Generation Computing Programme, the UK Alvey Programme, the European Esprit programme and, in the US, the DARPA Strategic Computing programme (these are further discussed in 3.1 and 7.2). Most of these programmes embodied what was then a fairly new emphasis on promoting industrial-academic links (Senker and Faulkner, 1995; Owen and Oakley, 1989) They were therefore much more than research-funding programmes, but provided an impetus to the AI industry, through the promotion of academic-industrial collaboration (cf 7.1). The identification of AI as a strategic technology was therefore important to the early success of the AI industry. The story of how this identification was made has been told by pointing to the institutional basis of AI in prestigious academic institutions (Fleck, 1982; 1987). However it is difficult to divorce the idea of institutional authority from the rhetorical and discursive practices that maintain that authority.

The question of how AI was identified as a strategic technology by policy makers (and as a growth market by industry and by venture capitalists) gained a more practical urgency, at least for some members of the AI community, in the light of the alleged 'failure' of AI by the late 1980s and early 1990s. I begin my case study (in Chapter Three) with a study of the way in which AI vendors construe and explain the failure of AI and how this partly depends on a construction of the history of AI as one in which a powerful technology was oversold. This is an interesting starting point because it enables me to explore how the AI industry construes its own identity through its relationship with its customers. Most interestingly, however,

---

[38] I shall not try to further explicate this term at this point. Cf discussion below, especially chapters Three and Seven.

this performed identity appears to include analyses of the failure of AI in terms of a failure in communication (between vendors and the market, and between academics and industry); thus it is an opportunity to explore how communication is adduced as significant to the transfer of technology by members of the industrial AI community.

In taking AI as a case study in technology transfer, I am deliberately distancing my study from the question of whether machines can think, and whether knowledge is inherently social, and similar issues which have usually been the focus of sociological studies of AI. Collins (1987), for example, suggests that AI researchers and sociologists of knowledge share a common subject, *knowledge*;[39] and he offers a typology of knowledge as a contribution to the common discussion.[40] In his later book, *Artificial Experts* (1990), Collins rejects approaching the discussion through a typology of knowledge.[41] Instead, he offers a re-reading of the question of machine intelligence in terms of social interaction (p 13). The issue is still taken to be the claim that machines can think, and Collins develops an alternative description of knowledge based in human action. Woolgar (1985; 1987) in a response to critics of AI such as Coulter (1983) argues for the retention of the sceptical stance and sees the AI debate as a contest between alternative essentialist accounts of what it is to be human. While this study distances itself from the substantive debate, Woolgar nonetheless takes it that AI is about building intelligent machines. Against all these, may be put the suggestion of the philosopher Hilary Putnam (1988, p 277) that, while AI researchers tend to believe in machine intelligence, the building of intelligent machines is at best a notional aim of AI research. In terms of my ethnographic perspective (cf 2.3) my starting point is similar to Putnam's. Methodologically,

---

[39] Collins (1987, p 329) says: 'Self-imposed evaluative neutrality is less easy for the analyst to maintain in the case of AI than in other areas of science because both AI and recent science studies share a topic: knowledge. This means that the findings of the whole research program of modern science studies, as well as specific studies of AI, have implications for what AI researchers are trying to do, whereas for most science and technology the visitor from science studies can be an onlooker, in this case he or she is also an expert, a knowledge specialist.'

[40] Interestingly, Collins' typology had included the suggestion that knowledge moves from the 'tacit' to 'the explicit', which is directly contrary to the suggestion made by Dreyfus and Dreyfus (1986) that only novices make use of rules. Collins, like several of the texts referred to in Chapter One of this thesis, makes use of Polyani's (1967) concept of 'tacit knowledge' (including Nelson and Winter, 1982; Faulkner and Senker, 1995). 'Tacit knowledge' is generally treated as knowledge which happens not to be made explicit. The phenomenological notion of 'embodied knowledge' by contrast, which is presumably informing Dreyfus' typology, is of a knowledge that is not available to explicit intellectualisation (cf Merleau Ponty's *Phenomenology of Perception*, 1945).

[41] Taking Dreyfus' approach to the typology of knowledge as his example, Collins (1990, p 21) suggests that such typologies tend to lead to a dichotomy between scientific and practical knowledge, which creates 'too much discontinuity between classes of knowledge'.

however, the question of how or whether ascriptions of 'intelligence' entered into the communication of the identity of AI is part of what I am investigating.

2.3 Ethnographic standpoint and the availability of material

The material I am drawing on for this thesis includes textual and interview material that I list in more detail later (section 2.4). My starting point in collecting material, however, has been my experience as a trade journalist specialising in the field of AI for nearly ten years (1984-1993). This covers most of the period of the rise and fall of the AI market (which I discuss in some detail in 3.1). Over this time, I interviewed representatives of practically all the industrial companies active in the AI industry in the UK, as well as many from the USA and France (and some other European countries).[42] I also interviewed many AI academics (again, mainly from the UK but also some from the USA, France and other places). I attended most of the major UK AI conferences (particularly the industrially oriented ones) and several in the US and France. As the editor (and main author) of a monthly AI newsletter (*Machine Intelligence News*), I was dealing from day to day with press releases concerning AI products and companies (which were the bread-and-butter matters of the newsletter), as well as other relevant topics such as national and international funding programmes. From 1992, when I became publisher of the newsletter, I also became a member of the AI Vendors' Association (since my newsletter was itself one of the service products associated with the AI industry). This has given me a wealth of material, but also constitutes a significant methodological problem to do with ethnographic distance. In the next few paragraphs I explore some of the problems of 'native' or 'indigenous' ethnography (cf Fahim, 1982) and ask how they may apply in this case. I conclude this section by locating the material that I gathered as a journalist within the broader scope of material that I use in this thesis.

Ethnographic studies have traditionally been seen as involving a balance between the experience of the insider and the authority of the outsider, exemplified in the figure of

[42] As I discuss further in 3.1, the first and most important location of the AI industry was in the USA. However, both France and the UK had an early and important local AI industry and market. The most evident gap in my material is the experience of Japanese companies. The Japanese industry and market was also an early and significant one. However, Japanese companies did not tend to try to sell into European and North American markets (one exception being an effort by Hitachi in the late 1990s, which I discuss in 3.1; another being the ill-fated Prolog machine, launched to rival American Lisp workstations, which was removed from the market almost as soon as it was launched).

the 'participant observer'. This is sometimes stated as an opposition between subjectivity and objectivity. As Clifford and Marcus (1986, p 13) put it:

> Since Malinowski's time, the 'method' of participant-observation has enacted a delicate balance of subjectivity and objectivity. The ethnographer's personal experiences, especially those of participation and empathy, are recognised as central to the research process, but they are firmly restrained by the impersonal standards of observation and 'objective' distance.

However, the idea that objectivity is a characteristic of location, of being an outsider, may be challenged by stories which show that outsiders may simply and grossly misunderstand what is going on. This was shown, for example, by Soraya Altorki, a US-trained anthropologist who returned to her home (Saudi Arabia) and carried out a study of women like herself, from elite families. Altorki (1982) claims that only by being one of the women (as she anyway was)[43] could she gain access to material that provided evidence of the structural power of Saudi women in making marriages and so controlling inter-family links. She contrasts this (p 171) to the 'sociological myth' that Arab women are passive objects of exchange in marriage arrangements. The point here is not that the insider is in a position of epistemological privilege[44] , or that the Arab women's self-view cannot be challenged. Indeed, there may also be limitations on what an insider gets shown simply because she is an insider. Fahim and Helmer (1982, p xix) point out:

> [T]he local anthropologist may not be taken seriously by informants if he probes types of behaviour that informants view as commonly shared knowledge, such as marriage customs, or he may be considered intolerably crude in broaching other topics, such as sexual practices.

There is, they suggest, room for the 'rude foreigner' described by Elizabeth Colson (quoted p xix; cf Colson, 1982)[45] ,

> who would be able to crash through barriers and ask the kinds of questions that may not be appropriate. People are willing to respond since they realise that the

---

[43] Altorki (1982, p 169) does point out that she returned 'somewhat of a stranger to my own culture', and had to reassume the role expected of her.

[44] The point is more to do with access and acceptability than the foundations of knowledge. Altorki (1982, p 172) says: 'I believe that only a female native anthropologist can possibly have some easy access to the data needed, although I am prepared to admit the possibility that a foreign female anthropologist might, under most favourable field work conditions, also gain such access in the long run. But in how long a run?'

[45] Fahim and Helmer's paper is the introduction to a volume (Fahim, 1982) based on a symposium (sponsored by the Wenner-Gren Foundation for Anthropological Research, held in July 1978 in Burg Wartenstein, Austria). Elizabeth Colson is one of the contributors to the volume, but the comment about the 'rude foreigner' is not in this paper and was presumably made during discussion in the symposium.

> anthropologist, a sort of 'innocent child', does not know.

Where does this leave the questions of objectivity and ethnographic distance? The above discussion suggests that there are different benefits and dangers for 'insiders' and for 'outsiders', which require to be taken into account in any ethnographic field study. However, it offers no guarantees to either insiders or outsiders merely on the basis of their location. It leaves open the more general question of the possibility of ethnographic distance, and whether insiders may not still be at a disadvantage in interpreting their own culture. It is worth noting, from the example cited in the previous paragraph, that Altorki had spent time 'outside'. Indeed, it is only because she had been trained (in the West) as an anthropologist that she could address the question of the adequacy of Levi-Strauss's kinship structures as a description of the social power of Saudi women. The question, otherwise, would simply not have arisen. This is not to smuggle anthropology back in again as a source of knowledge, but to reintroduce it as a possible source of questions, hypotheses and interpretations.

If objectivity is not guaranteed by the poles of location ('inside' or 'outside'), is it to be found at the boundary? Crapanzo (1986), in an essay that accuses anthropologists of bad faith in so far as they believe they can be objective (the 'ethnographer's paradox'), points to the significance of boundary-making in attempts to understand another culture. He says (p 52):

> The ethnographer ... marks a boundary: his ethnography declares the limits of his and his readers' culture. It also attests to his - and his culture's - interpretive power.

In a similar point, Woolgar (1988b) describes ethnographers as crossers of boundaries. Again, this is in the context of no longer looking for objectivity, but retaining a search for the source of ethnographic distance. The question then becomes, how easy is it for the insider to mark and maintain boundaries?[46] The question of marking or maintaining boundaries is particularly pertinent in my own case since, unlike Altorki, I did not approach the field of AI as a trained

[46] Geertz (1983, p 151) observes: 'We are all natives now, and everybody else not immediately one of us is an exotic. What looked once to be a matter of finding out whether savages could distinguish fact from fancy now looks to be a matter of finding out how others, across the sea or down the corridor, organise their significative world.' Most people, however, do not address the question of how significative worlds are organised. We may all be natives, but we are not all anthropologists (either indigenous or visiting).

ethnographer. Altorki had available a contrast between 'being an anthropologist' and 'being one of the women', performed through an understanding of the difference between questions which were relevant to each community.[47] For example, as an anthropologist, questions of the likely benefits of different marriage links were only relevant (only discussed in her paper) insofar as they were questions that were decided by the women; she may have had strong views on who should marry whom, but we see no trace of this in her paper.[48] On the other hand, although I may not have had ethnographic training at the time, I came to the field of AI as an outsider, in two senses: I was a journalist and I had (previously) trained as a philosopher. However, if ethnographers are crossers of boundaries it does not follow that all crossers of boundaries are ethnographers. Both these backgrounds contributed different ways of constituting myself as different, and highlighted slightly different problems in my attempt retrospectively to recast ten years of my life as an ethnographic field trip.

As a journalist, I began specialising in the field of AI in 1984 when I worked on an engineering trade journal[49] and I began to focus exclusively on the field in 1985, when I became editor of an AI newsletter called *Machine Intelligence News* (MIN). This newsletter covered the AI[50] industry, including product announcements, company news, the development and implementation of applications, government initiatives, and the views of various market and strategic analysts. It was sold by subscription only and the readership (which was never more than a few hundred) consisted of people already with some interest in AI, including members of corporate AI R&D departments, market analysts, and AI vendors and consultants. Because the newsletter was about and for the industrial AI community, this minimised the sort of

---

[47] I am not suggesting that the availability of this contrast solves the problem of marking boundaries in Altorki's case, merely that it is one of resources available for managing the problem. Altorki's point, however, is to cast doubt on the very notion of 'indigenous anthropology'. She says (1982, p 174): 'At no point in the analysis of my data have I found that the methods and theoretical perspectives developed in social anthropology were inadequate for an ethnological, cross-culturally meaningful, interpretation of the position of women in my own society.'

[48] This sort of distinction is deliberately blurred by anthropologists concerned with the meeting of autobiography and ethnography (cf papers in Clifford and Marcus, 1986; Okely and Callaway, 1992).

[49] *Engineering Today*, published by Haymarket. In the three years I was there, this successively changed its name to *Technology*, then *New Technology*, before being closed down in 1985.

[50] Until about 1987, MIN covered the AI industry in the terms described in the first paragraph of this section, which might otherwise be described as 'symbolic AI' (I discuss this in Chapter Five). Later in the 1980s, when products and applications of artificial neural nets began to appear, I also covered that. By the late 1980s, I extended the coverage over new interface technologies such as virtual reality.

journalistic boundary crossing that might have been involved if I had been writing for a publication with a more general readership.[51] I did not, for example, provide parenthetical explanations when I used technical or insider terms, as would be necessary in a more general publication.[52] On the other hand, I performed a boundary-crossing role within the industrial AI community, providing a medium of communication between sub-communities. This was evident, for example, when I interviewed AI vendors who spoke to me in order to communicate with my readers who were their potential customers, while I attempted to gather the information that I thought my readers would want (I discuss some examples of this in 3.2). Despite these internal divisions however, the readership can be seen as a performed community, and the issues and questions covered by MIN reflected this community. Relevant (and pressing) questions included the extent to which any 'real life' applications of AI had been or were about to be implemented, whether and when the market was going to take off and, finally, what went wrong (why did the AI market fail?). As a journalist my interests (in the sense of questions I understood to be relevant) largely coincided with the interests of the industrial AI community.

Perhaps my philosophical identity can introduce a sceptical difference? As a philosopher, I came to the field of AI with an interest in what I took to be the philosophical issues raised by AI, in particular its claims about intelligence.[53] However, these turned out to be at best marginally relevant to the newsletter.[54] In this respect, I am like the indigenous anthropologist who sees how much outsider anthropologists misunderstand (cf my comments on the sociological literature on AI in 2.2 above). Nonetheless, like Altorki (1982, p 174), I do not see a problem in principle in applying my more newly learned sociological skills to material gathered

---

[51] One example of this was when I tried to interest *New Scientist* in a story about the first 'real-life' implementation of neural networks (in an airport security device). For readers of MIN, the question of whether, how many, and how successfully AI programs were going into implementation was a major story throughout the late 1980s. But it was not a story for *New Scientist*.

[52] On journalist training courses we were taught that parenthetical explanations should always be provided as if the reader had no specialised knowledge, and more knowledgeable readers would simply read over or ignore the parenthetical explanations. To have included such parenthetical explanations in MIN, however, would have distanced it from its readers. Similarly *Engineering Today* would never have explained engineering terms that might have required explanation in a national paper. This is one of the ways in which the trade press performs community.

[53] I published one academic paper (Vaux, 1986) in which I argued against the AI conception of mind, but also that AI and philosophical contributions to the AI debate were not engaging in the same issues. A much developed version of the latter argument is contained in Chapter Six (section 6.1) of this thesis.

[54] I occasionally included brief reviews of new contributions to the AI debate, but I always felt this to be a bit of a self-indulgence. My connections with the journal *AI and Society* gave me some context for continuing to explore these interests.

while I was a journalist. Also like Altorki, however, I need to point out some specific practical issues in relation to the gathering of material.

As a journalist, I spent a great deal of time interviewing various significant figures in the industrial AI community, particularly senior executives in AI start-up companies. I tape recorded these interviews (as I would a sociological interview), but in general I neither transcribed them fully nor kept them for any length of time after I had extracted what I wanted for the immediate purposes of a story. On journalist training courses we were taught to keep notes and other material for six months in case of libel charges. As a philosopher, I had a tendency to notice arguments that reflected philosophical issues and a sensitivity to anti-philosopher jokes.[55] However, philosophers have no concept of observational material, and I kept no field notes of any of these conversations (although I had in mind, for most of the time, the idea of 'one day writing all this up'). Some of this material, therefore, is not documented and may not be reliable in detail, although it has as much reliability as some other published autobiographical material. I therefore make use of some of my undocumented memories, but always make it clear (eg by using such forms as 'I remember ...'). However, since I was writing a monthly newsletter from 1985 to 1993, I do have a great deal of textual evidence of the stories of moment and how they were handled. The final year of the newsletter overlapped with the first year of writing this thesis, and for this period I do have much better documented material, including interview transcripts and field notes of a Prolog users' conference (I make use of this material particularly in Chapter Three). In addition to the material I gathered as a journalist, much of this thesis is based on written material (published and unpublished) which provides examples of explanations of AI in different contexts and to different audiences. This is the major documented base of the larger part of this thesis.

2.4 Summary: explanatory practices

In this chapter I have explored what it might mean to study technology transfer as communication between discursive communities, in terms of the methodological traditions of STS (section 2.1) and some of the specific technology transfer issues

[55] For example: 'Well philosophers haven't got very far with this subject in two thousand years.'

raised by the case of AI (section 2.2). Discussion of the methodological tradition provided a notion of 'performed community', an idea which was also raised in the slightly different context of the question of my own ethnographic position (section 2.3). In this section, I try to draw some of these issues together and to ask what the lessons are for this case study. In particular: what are the discursive objects of my case study and how should I approach them? The point at issue is communication *between* discursive communities, and therefore it is not enough simply to explore the performance of a single community. I need not only to look at the industrial AI community, but also at the academic AI community (and probably several other communities besides). How does each construe their own identity and the identity of AI; and how (since the issue is communication) do they construe one another? If I focus on how each performs community, however, there is a danger of making it seem as if communication is impossible. To show the negotiation of a 'we' and of an 'other', is to show the negotiation of boundaries and of difference. The point, however, is to show how communication is possible, as well as the difficulties involved. In particular, since this is a study of technology transfer, the point is to show how communication about science and technology (what it is, what it is for, how it may change things) is possible. The solution, therefore, is to look at how scientific knowledge or technological advances are described and explained by different communities, particularly how they are described between communities, and then to read these descriptions for the performance of community. In the next few paragraphs I discuss some of the methodological implications of this.

It is important, first, to note a problem in using the phrase 'technology transfer', which is often taken to imply a hierarchical distinction between science and technology, and a unidirectional, linear model in which science is done in universities, transformed by industry into technological products, and then disseminated (cf 1.2). Many authors have introduced other terms in which to conceptualise the knowledge used in innovation (cf Faulkner and Senker, 1995, pp 213-227). Nonetheless it seems to me valuable to retain the phrase 'technology transfer' to refer to institutionalised political practices for supporting technological innovation at a national level, as well as corporate practices of identifying, evaluating and marketing innovatory processes and products. In this narrow sense, only Chapters Three and Seven of my case study deal directly with technology transfer. It is central to my

analysis, however, that technology transfer should be located within a broader context of practices of communication and explanation. Academic explanatory practices, addressing a variety of audiences, are particularly relevant in a case of technology transfer which was research-led and which, on the basis of academic research developments was claimed to have revolutionary social and economic implications - as was the case with AI. Chapters Four, Five and Six all address communication between discursive communities, within the broader context of technology transfer, largely within and between academic disciplines and subfields, but also in relation to more general audiences.

I use the term 'explanatory practices' to refer to discursive practices of describing and explaining. My interest is in whether and how such explanations may perform more than they overtly say (eg descriptions which act to do more than simply describe). More specifically, I am interested in whether and how they perform community and configure the audience. It is important therefore to present any explanatory practices that I study in terms of *location*, *context* and *audience*. To give an initial indication of what I mean by these terms: by 'location' I mean the location of author or speaker or producer of the text, perhaps in institutional terms (for example, AI academic); by 'context' I mean the site of the explanation in a material sense, that is the type of publication (eg textbook) or discussion (eg interview) in the understanding that there are complex performed conventions appropriate to different contexts[56] ; and by 'audience' I mean both the intended and configured audience. The category of 'audience' is not straightforward, and I am using a distinction between 'intended' and 'configured' that needs some explanation. The intended audience, I take it, is the audience of publishers' contracts, or the audience the author personally wishes to influence or to produce. A configured audience, on the other hand, has to do with mechanisms of capture. This is largely following the description given by Woolgar (Cooper and Woolgar, 1993; Grint and Woolgar, 1997; Woolgar,1991), where the user is configured partly through internal textual mechanisms and partly by various commentary resources (marketing material, reviews, visions of the inventor, etc). However, it is not clear what the equivalent of commentary resources would be in the case of explanatory practices, since relevant

[56] However, there are are further senses of 'context' which may also be relevant, including, for example, historical context (eg the way computers were talked about in the 1950s), and programmatic contexts (eg the context of the 'need' for a national IT programme).

commentary resources would themselves be explanatory practices.[57] It should not be assumed that any particular location, context or audience can be determinately defined, since these identifications will always be negotiated outcomes and as such a potential object of study. Nonetheless, there are some institutionalised and relatively stable categories at issue here (such as 'sociologists' and 'textbooks'), and I assume it will usually be relatively easy to make a preliminary identification.

A study of explanatory practices has a number of immediate advantages in tackling a case study of the exploitation of AI. First, as an academic field, AI has been explaining itself in different contexts to different audiences for several decades. Much of this material is surprisingly available, partly because this is a community with a strong sense of its own history. Even the sort of material that might in other case studies have to be sought out in archives is available on the net (for example, the text of the funding proposal for a summer school held in Dartmouth in 1956, often described as 'the first AI conference', is published on the web page of the AI researcher who was its main author, John McCarthy). Secondly, a public debate with academics from other disciplines has been ongoing also for several decades. This debate allows an analytic switch of location, to study explanations and descriptions of AI from other disciplinary locations, and also provides one sort of example of communication between different discursive communities, and a study of contested identifications. Thirdly, the project to launch AI on the market was both discussed and carried out at least partly in papers and seminars directed at industrial researchers (cf Chapter Seven). In addition, the identification of AI as a strategic technology and as the future of computing has left at least some traces in publicly available material, including programmatic lobbying (eg Feigenbaum and McCorduck's (1984) *The Fifth Generation*), committee discussions (eg the report of the Alvey Committee (DOI, 1982)), and histories told by interested parties (for example Brian Oakley's post mortem on Alvey (Oakley and Owen, 1989)). Finally, stories of the failure of AI, and what went wrong, were told by many in the industrial AI community in the early 1990s (and this is available to me in material that I gathered as a journalist). These stories not only provide a further location from which the identity of AI is told, but adduce an explanation of the failure as a failure in

[57] That is, if an explanation does not convince further explanation may be required, but not usually explanation of the explanation. This may be a general issue when the metaphor of 'technology as text' is reapplied to texts.

communication, and specifically as a failure in explaining and describing AI. This explanation for the failure of AI also provides a narrative starting point, not so much as an hypothesis to be tested, but as a recurrent response to explanations and descriptions of AI. That is, the question is not whether the diagnosis (overselling) is right but: how do accusations of 'hype' act as a performance of community on the part of audiences to descriptions of AI, when it comes to their turn to speak?

The structure of the case study should be noted. Historically, I begin at the end, with a study of the AI industry of the 1980s and early 1990s. The reason for this is to begin with the diagnosis of a problem (that hype caused the AI market to fail) which contains an implicit, if crude, theory of technology transfer, which I use as a narrative resource within my case study. The following chapters, starting with a study of the early days of the academic AI community from the 1950s (Chapter Four) are more or less in historical order, although there is some overlap because the primary structure is thematic, dealing first with historical narratives (Chapters Four and Five), then with reading and interpretation (Chapter Six), before returning to the implications for the industrial and political exploitation of AI in the early 1980s (Chapter Seven). Throughout the case study, and as a matter of analytic principle, I am tracing the explanatory practices associated with AI, at different locations, different times, in different contexts and to different audiences. This is not intended merely as an historical journey through different usages, but is meant to pose the question of how explanations of AI differ across various locations, contexts and audiences, and to be able to pose the question how this history of usages was apparent in the strategic and tactical judgements made about the potential of AI in the early 1980s. The material is largely text-based (except for interviews that I carried out as a journalist). I have not carried out further interviews with members of the AI community during the writing of the case study. Although this would have added a further historical dimension it would have yielded material that was not performative within the marketing of AI (that is, the interviews would have provided reminiscences to a sociologist). While this would have been interesting and relevant, I could not justify it within the length and time allowed for writing a PhD thesis. The chapters in the remainder of this thesis are organised as follows:

Chapter Three - How the AI industry construed the problem of AI. This chapter is based on descriptions of AI located within the industrial AI community. I ask about the relationship between 'indigenous commentators' (eg analysts, journalists, corporate marketing departments) and the interests of the industrial AI community, and how these interests were performed through the production of market predictions and analyses of news events. Finally, I explore the terms of an adduced 'failure' of AI, and various diagnoses of the problem given by members of the industrial AI community, and ask about the implications for a theory of technology transfer.

The following chapters (Four to Seven) explore how AI was described in different contexts and to different audiences in order to understand the socio-rhetorical mechanisms involved in communication between discursive communities.

Chapter Four - Narratives and audience: AI acquires a history

In this chapter I ask how histories of AI may vary according to audience. I focus on the early history of AI (from the 1950s to the 1970s), both as it has been told in histories for a general reader and as the field was explicated in peer discussions and textbooks. I ask whether the configured audience may be understood as socio-rhetorical mechanism determining differences in the public and peer histories.

Chapter Five - United around the symbolic? Alliances in the field

This chapter addresses the performativity of history, its power, for example, to cement alliances or overturn a previous order. I draw on two examples: first, the history of cognitive science as an alliance of disciplines; secondly, the rehabilitation of the 'neural net' or 'connectionist' paradigm at the expense of 'symbolic AI'. I conclude by asking how an understanding of the way in which historical texts may perform to further the interests of a faction, may throw light on communication between discursive communities.

Chapter Six - Reading and interpretation: ascribing intelligence

In this chapter I turn from questions concerning the performative power of texts, to asking about the interpretive power of the reader. I do this partly by exploring differences between disciplinary readings of issues, in the context of the AI debate,

and ask whether this can generate a concept of disciplinary 'interests'. I then ask how the interpretive flexibility of ascriptions of intelligence may relate to attempts to control and configure the reader.

Chapter Seven - Industrial and political strategies: AI as strategic technology

This is the final chapter of the case study, and in it I look at explanations of AI within the context of attempts to transfer the technology to industrial use. I look at differences in explanations for a heterogeneous industrial audience (consisting both of researchers and of corporate executives) and in the explanations given within policy making discourse. I ask in each case how attempts are made to construe the need for the technology for a specific audience. In drawing the case study together, I also ask whether and how the trope of intelligence was deployed in explaining AI to these audiences, and finally whether and how AI may be said to have failed (the question raised in Chapter Three).

Chapter Eight - Analytic and theoretical conclusions

This chapter draws together the major themes in the case study and offers a reassessment of the methodological and theoretical issues raised in Chapters One and Two, as well as identifying new methodological and theoretical issues arising in the course of the case study. I conclude by asking how the discussion in this thesis may be generalised, and what avenues for further research are suggested.

CHAPTER THREE - HOW THE AI INDUSTRY CONSTRUED THE PROBLEM OF AI

3.0 Introduction.

This chapter has a two-fold purpose. The first is to introduce the 'problem of AI', that is, a problem as adduced by members of the industrial AI community during the late 1980s and early 1990s. This was perceived as a problem specifically of technology transfer, and one popular explanations of the problem was that overselling or 'hype' in the early days of the AI industry had led to unrealistic market expectations and therefore caused a collapse of the market. In this chapter, I show how discussions within the industrial AI community produced this explanation. I then argue (in 3.4) that this diagnosis may usefully act as a sort of narrative motif to set against my own working hypothesis that technology transfer is to be seen as communication between discursive communities, and that it involves the production of social relations between those communities. The second purpose of this chapter is to provide a study of the industrial AI community of the 1980s, looking at the explanatory practices of the industrial AI community, and whether and how these relate to location, context and audience as described in my methodological aims (section 2.4). This study will take its place within a series of historical studies (through the succeeding chapters of the thesis), looking at the explanatory practices of different discursive communities, including the academic AI community (Chapters Four, Five and Six), other disciplinary communities, including Cognitive Science (Chapter Five), connectionism (Chapter Five) and philosophy (Chapter Six), and corporate and policy-making communities (Chapter Seven). My purpose is to look at the different explanations given by these different communities, and also to look at any differences within a discursive community's explanatory practices that can be seen to be related to context and audience, to ask whether any common or neutral explanation of AI may be identified, since knowledge-passing models of technology transfer would seem to imply such neutral explanations should be identifiable. The double purpose of this chapter is made easier insofar as the 'problem of AI', its failure to attain the markets originally predicted for it, was a dominant concern for the AI industry through most of the 1980s.

In accordance with the methodological approach that I described in 2.4, I start this section with an indication of the location, context and audience of these explanations and descriptions. One major source of the material used in this chapter was the specialist commentators, including my own newsletter and other AI newsletters, magazines and specialist market reports. The AI newsletters (like other specialist industrial newsletters) were not particularly visible to outsiders. They were not available in newsagents, and were usually sold by mail order subscription only, relying on a relatively high subscription rate to service a relatively small subscriber base,[58] and they were less visible even than the mainstream computer trade press which is usually 'controlled circulation'.[59] Among AI newsletters and subscription-only magazines available in the 1980s were:

US

Expert Systems Strategies (later *Intelligent Software Strategies*). US Monthly newsletter. Started 1985

ICS Applied Artificial Intelligence Reporter (later *AI Week*; later *Intelligent Systems Report*). US. Monthly Newsletter. Started 1983

The Spang Robinson Report. US. Monthly newsletter. 1984-1992.

UK

AIBusiness. UK. Monthly newsletter. 1986 - 1989. (AIBusiness shared a publisher with *Expert Systems User* and in 1989 the two were combined in *Expert Systems User*)

AI Watch. UK. Monthly newsletter. Started 1992 (still published).

Expert Systems. UK. Quarterly magazine. Started 1984

Expert Systems User. UK. Monthly magazine started 1984 (newsletter from 1989) Now defunct.

Machine Intelligence News. UK. Monthly newsletter. 1984-1993.

France

Lettre d'IA. France. Monthly Newsletter. Started 1985.

This is not an exhaustive list: there are at least some US newsletters left off the list. I have not included any academic AI journals or publications of professional bodies associated with AI as I do not draw on them in this chapter. I have also drawn (especially in 3.1) on market reports by the UK market analyst Ovum. While several market research companies (eg Frost & Sullivan) ventured into analyses of the AI industry during the 1980s, Ovum began by specialising in the area. The company's first report (Ovum 1984)[60] , written by founder Tim Johnson, was *The*

[58] For example *MIN*'s circulation was in the hundreds; an annual subscription was £110 in 1987, and £170 in 1993

[59] A controlled circulation newspaper is sent free to computer (or other relevant) professionals and makes its money mainly from job advertisements.

[60] The Ovum reports were written by a number of different authors. For clarity I have treated Ovum as the author in bibliographical references; individual authors are named in the Bibliography.

Commercial Applications of Expert System Technology, and Ovum produced a number of reports on expert systems during the 1980s (while also expanding both in numbers of staff and the technological areas covered in its reports). Other material is taken both from the newsletter I edited (*MIN*) and from other AI specialist newsletters.

The texts described above, on which I have drawn in this chapter, may be said to be located in a community of indigenous commentators. The audience for these texts is the AI industry in a broad sense, that is, not only vendors but also other analysts, investors and early customers. It may seem counter-intuitive to place customers 'inside' the industry. However, what Ovum (1986a) calls 'leading users', the companies that were buying AI products during the 1980s, were mainly buying development systems in order to explore and build AI systems either for inhouse use or to sell as products. That is, customers at this time tended themselves to be AI developers. The indigenous commentators, then, were writing about and for an insider audience. Not only were the commentators 'insiders' and in some sense shared the interests of the broad industrial AI community, but they also formed a specific subcommunity whose apparent role was to facilitate communication between other specific subcommunities. For example, when I was interviewing AI vendors, most of what they said to me was directed at my readership which (for much of the 1980s at least) was located in their main market. Such interviews are often highly managed, with other corporate representatives in the room, including someone with technical knowledge, and a PR person.

In addition to drawing on published material, such as newsletters and market reports, I have also drawn on other material that I gathered when working as a journalist. This includes some material from taped interviews beyond the material that I published at the time. My presentation of the material is also informed by my memories of discussions, and occasionally I present these recollections as material (but always indicating that they are personal recollections). Towards the end of the period, I also made field notes of discussions and conversations at a Prolog users conference which I have drawn on especially in section 3.3.

This chapter begins (3.1) by asking how the AI industry and AI market were

identified and described by members, through a study of a series of reports from one market research company (Ovum), asking how the interests of the industrial AI community were negotiated through the reports, and what this example indicates about the situation of indigenous commentators. In the following section (3.2) I ask how the AI newsletters represented the strategies adopted by AI vendors to manage the perceived problem of the AI marketplace; again this section involves looking at the relationship between indigenous commentators and the community of which they are a part. I then (in 3.3) explore some different explanations of the problem of AI, as given by members of different subcommunities within the broad industrial AI community. Finally (3.4) I ask what is implied for a theory of technology transfer and how I plan to address this within my case study.

3.1 The AI market: identifying growth and crisis

This section provides a brief sketch of the AI industry and its market as a preliminary to looking at ways in which AI vendors attempted to manage their situation in the market (cf 3.2). Of course, to identify a market sector is already a political act.[61] In this section I look at how the industry and its market was construed by members, taking as my example a series of reports written by the UK-based market research company Ovum (described in 3.0). I begin with a sketch of the AI industry and its marketplace, and then look at some predictions of market growth and ways of explaining market failure, asking whether and how such predictions and explanations may themselves perform community.

The first AI vendor companies were set up, mainly by academics, in the late 1970s and early 1980s. They sold hardware and software products that had been developed (or initiated) in university research labs, mainly in a small number of high ranking academic institutions (Fleck, 1982; Ovum, 1984). Many of these products were targeted at the development of expert systems (a number of prototype expert systems had already been developed, mainly as PhD research projects) (cf Minsky, 1968, p v); and they were primarily written in AI programming languages such as Lisp or (in Britain and France) Prolog. The products in the early stages included

[61] Cf Henkel et al (1999) for a description of the Materials Panel of the UK Foresight initiative as attempting to renegotiate the DTI's definition of the materials industry.

expert system development environments (shells), some bespoke expert systems, AI programming languages (Lisp or Prolog), and Lisp workstations (ie computer workstations optimised to run the Lisp language). The industry was largely based on small start-up companies. Table One shows some of the earliest small AI companies. As Table One indicates, most of the small companies involved in the first wave of commercialisation had links with university research groups, and most of the products were commercial versions of what had till then been research software or hardware.

TABLE ONE: THE FIRST WAVE OF SMALL COMPANY START-UPS

Yr Founded	*Name*	*Main products*	*Research/Univ*
(1979)	Inference ART	(ES shell)/1983	(Cal. Tech)*
1979	Intelligent Terminals	(ES shells/rule induction)	Edinburgh
1980	Lisp Machine Inc Lambda	(Lisp workstation)	MIT
1980	Symbolics Inc	(Lisp workstations)	MIT
1980	IntelliCorp	KEE (ES shell)	Stanford
1980	LPA	Micro-Prolog (prog lang)	Imperial
1981	Teknowledge	S.1, M.1 (ES shells)	Stanford
1981	Expert Systems Ltd	Prolog-1 (prog lang)	(Edinburgh)†
1981	APEX	(ES for financial applications)	--
1983	Carnegie Group	XCON (for DEC)	CMU
		SRL (ES shell)	

* Inference was founded in 1979 as Systems Cognition Corp, selling a maths package called SMP developed at the California Institute of Technology; in 1983 Inference launched its Lisp-based expert system development tool, ART, which had originally been developed for inhouse use (Ovum 1984, 277-8).

† ESL was not a university spin-off, but its Prolog was the standard version of the language, developed at Edinburgh.

In several cases, leading AI academics were involved in founding the companies: for example, Ed Feigenbaum (Stanford) was a founder both of IntelliCorp and Teknowledge; Donald Michie (Edinburgh) founded Intelligent Terminals; Mark Fox, John McDermott, Raj Reddy and Jaime Carbonell (all of CMU) were on the founding board of the Carnegie Group; Keith Clark and Frank McCabe (Imperial) founded LPA (Logic Programming Associates); and a research team at MIT building the CADR Lisp machine split into two rival Lisp machine companies - Russell Noftsker leading the majority into Symbolics, and Richard Greenblatt and others joining LMI. Given that the academic AI community was small and interlinked and

predominantly North American (Fleck, 1982), the products had a certain commonality and were predominantly Lisp-based.[62] From 1984, the number of AI companies proliferated. Among the 1984 start-ups in the US were several companies offering Lisp programming environments, including Gold Hill, Franz Lisp and Lucid. Quintus, which became the leading supplier of Prolog was also incorporated in the US in the same year. Larger corporations which entered the market during the early 1980s included: Xerox which produced a Lisp workstation based on technology developed at Stanford University; and Texas Instruments and Sperry[63] both of which sold rebadged versions of LMI's Lambda machines (called Explorer). DEC's VAX workstations also became popular AI development platforms, and in 1984 DEC launched a version of Lisp to run on VAX. By 1986 a two-volume report published by market analyst Ovum counted 22 software suppliers and nine hardware suppliers in North America, and 35 software suppliers and 11 hardware suppliers in Europe (Ovum 1986a and 1986b).[64] A report two years later counted a world total of 130 vendors offering 211 products (Ovum 1988).

The small AI industry was given a boost by the institution of a number of national and international Fifth Generation funding programmes, first in Japan and then elsewhere, as shown in Table Two. These funding programmes were significant for prioritising AI research within national science funding budgets. The market for AI in the early part of the 1980s was largely made up of systems sold to R&D teams in university and corporate labs for technology evaluation and the development of experimental and prototype systems. The fifth generation programmes meant that

---

[62] The smaller European AI community tended to work in Prolog (a language developed by researchers in France, Britain and Hungary), which was also the language adopted by the Japanese Fifth Generation Computing Programme announced in 1982. Some of the earliest applications of Prolog were made in Hungary. A 1986 briefing document produced by Balint Domolki, director of the Hungarian Computer Research and Innovation Centre, SZKI, claimed 'Prolog arrived [in Hungary] in 1975, as elsewhere through academic channels, but the environment in which Prolog was used happened not to be an academic one. .... [B]y the mid-1970s Prolog was used in Hungary in several fields of computer applications including pharmaceutical research and architectural design. Some of these programs are now regarded as the first European achievements in 'expert systems' (a notion which did not yet exist at that time).' More singular approaches included the rule-induction programming that Michie's company offered which was based on an algorithm originally developed by an Australian, Ross Quinlan.

[63] In 1986 Burroughs acquired Sperry to form Unisys which continued to market Explorer.

[64] This involves counting some of the hardware suppliers twice, since US hardware suppliers also figure in the European list, alongside Bull, ICL, Nixdorf and Siemens. The European list of software suppliers includes R&D institutes, whereas the North American list is confined solely to corporate suppliers of commercial packages.

university AI teams had a budget for AI products (hardware and/or software).[65]

TABLE TWO: FIFTH GENERATION FUNDING PROGRAMMES

Yr started	*Name*	*Country/Org'n*	*Value/length**
1982	Fifth Generation Computer Project	Japan/MITI	Y100,000m/10 yrs†
1983	The Alvey Programme for Advanced Information Technology	UK	£200m/5 yrs (+£150m from industry)
1983	Strategic Computing Program	US/DARPA	$600m/5 yrs~
1984	Esprit (European Strategic Programme for R&D in IT)	Europe	1500m ECU/10 yrs**

*Value and length as originally announced.
†Moto-oka & Kitsuregawa, 1985; the Alvey Committee (DOI, 1982, par 3.18) translated the value of the Japanese programme as between $200m and $500m. ~DARPA, 1983(b).
**This was the May 1983 proposal; the initial project was for 750M ECU over 4 years (1984-8) (cf House of Lords, 1984-5).

In what sense was there an AI industry, and what was its market taken to be? The Ovum reports largely addressed the market as one for 'expert systems' (or in the final report in 1989, for 'knowledge-based systems').[66] In its first report (1984) Ovum started by defining an expert system, in terms of likely uses (application areas) and its software architecture (ie knowledge base, inference engine and user interface). The 1984 report introduces expert systems by distinguishing between the 'few spectacular pioneering systems' associated with Ed Feigenbaum's group and Stanford University (including a geological exploration system called Prospector, a medical diagnostics system called Mycin, and a Dendral, a system for determining molecular structures) and a 'new generation of expert systems' offering 'cost-effective solutions' for 'day-to-'day problems'. Ovum commented (pp 1-2), on the latter:

> These systems are less spoken about in the media because they are in less glamorous applications than medical diagnosis or mineral prospecting. They are

[65] In the UK, the Alvey programme standardised on the Xerox Dandelion, a Lisp workstation that was considered more cost-effective than the top of the range, and more expensive, Symbolics and LMI machines.
[66] The reports I am drawing on are: *The Commercial Application of Expert Systems* (Ovum 1984); *Expert Systems* 1986 Vol 1: USA and Canada (Ovum 1986a); *Commercial Expert Systems in Europe* (Ovum 1986b); *Expert Systems in Banking and Securities* (1988a); *Expert Systems Markets and Suppliers* (Ovum 1988b); *Knowledge-based systems: Markets, suppliers and products* (Ovum 1989). Ovum also published reports in other AI related language such as natural language processing.

> carrying out such tasks as configuring computer systems, planning the repair of telephone cables, and helping people to use complex software. But it is in this more technical area that the best immediate prospects for expert systems lie.

In subsequent reports it is assumed that the reader probably knows what an expert system is (and any definitions are much briefer), but the question of most likely application areas continued to be an issue for predicting the market. In its 1986 report on North American markets, Ovum reported a database of 500 expert system projects, in which 'leading examples include fault diagnosis, analysis of complex data and advice on regulations' (1986a, 19). As the financial sector became increasingly promising, Ovum published a special report on *Expert Systems in Banking and Securities* (1988a).

The main purpose of the Ovum report, however, was market estimation and prediction. The figures that Ovum gives from one report to the next are difficult to compare by categories, because they use slightly different terms reflecting changes in the market. However, a pattern of assumed growth is evident from the following examples:[67]

- 1984: US market for expert systems development projected to be worth $95 million in 1985 and $685 million in 1990;
- 1986: North American product and services market for expert systems estimated at $400 million in 1986, and predicted to be worth $1.9 billion in 1992;
- 1988: total traded ES development and delivery market in the US estimated at $587m in 1987 and predicted to be $2.8bn in 1992 ($5.8bn in 1995).

In providing market figures, the Ovum reports also supplied a commentary on the marketplace, and reading this commentary provides a history of some of the changing concerns of the AI industry. In 1984, for example, it is explicitly assumed, first, that growth of the market will be limited by supply rather than demand (because 'expert systems will be able to deliver such clear commercial benefits' (p 36)) and secondly that the main bottleneck will be the availability of skilled knowledge engineers. In 1986, the report asked why there were not more operational systems (that is, real rather than prototype systems), and commented sceptically (1986a, p 2):

> One common reply is that commercial companies who achieve operational systems will not be keen to advertise this fact, but will prefer to keep secret a system that

[67] The figures I give here do not reflect the detail of figures given in the reports.

> perhaps provides a competitive edge. In fact our research suggests that this is not usually the case.

Nonetheless it reported that 'leading users' expected soon to implement expert systems and expected to derive substantial benefits. In 1988, however, the Ovum report bluntly opened with the words (1988b, p 1):

> The expert systems business has been through a terrible period in 1987-88. In the USA, the 'Gang of Four'[68] - the new venture software companies which once led the industry - have all run into losses and cut back staff more or less severely.

After years of growth, many expert systems companies had run into reversal and losses. Nonetheless Ovum (pp 2-3) confidently projected a 'second wave' of growth, based on three drivers: growing numbers of operational applications; a mature technology, 'much closer to the real needs of mainstream data processing'; and 'a more realistic marketing focus' . Finally, in 1989 (p 3), Ovum continued to perceive an increase in the number of operational systems, a better defined technology and a more mature supply industry and concluded that the knowledge-based systems (KBS) business was in 'much better health than it enjoyed only a year ago'. It also noted 'a change of style', in particular a tendency to use the phrase 'knowledge-based system' instead of 'expert system'. Ovum reflected this change in the name of its 1989 report, commenting (p 3):

> 'Knowledge-based systems' has the advantage of identifying this type of computer system by the technology used - search and inference from a knowledge base - rather than giving the misleading impression that it is the attempt to emulate an expert which is the key distinguishing factor.

The above passage presents the renaming of expert systems[69] as 'knowledge based systems' as a reassertion of the technological. The authoritative tone of these remarks belies the sense of crisis in the industry. Between 1984 and and 1989 the Ovum reports had moved from the expectation that the benefits of expert systems would drive a strong market demand (1984), to worrying about the slow delivery of operational systems (1986a), to reports of a crisis (1988b), to the description of an industry attempting to redefine itself (1989). The latter point is symbolised through

---

[68] That is Intellicorp, Inference, Teknowledge and Carnegie Group. This use of the term 'Gang of Four', without explanation, acts as a test of membership for the reader which is perhaps not appropriate in a market report, even one for insiders. The 1989 report, however, did spell out the names (p 8).

[69] For further comments on this renaming cf section 3.3. It is also worth noting that the name 'expert systems' was rejected in discussions leading to the setting up of the UK Alvey Committee in the early 1980s (cf section 7.2) in favour of 'Intelligent Knowledge Based Systems' (IKBS).

a change of phraseology. In both noting and deploying this change of phraseology, the Ovum report may be seen as intending to perform community; that is to say, the use of 'insider' terminology acts to say 'I am a member of the community'; and use of insider terminology also acts to maintain a discursive community. At the same time, the continuing predictions that growth would still occur may, as the 1988 report itself implies, also be seen as performing membership of the industry (1988b, p 1):

> The pessimist can match the current commercial difficulties with corresponding scepticism about the prospects for expert systems as a worthwhile and distinctive technology, at least as far as the mainstream dp business is concerned. [....] Not surprisingly, Ovum takes a more positive view.

The phrase 'not surprisingly' reinforces the performance of membership, configuring an audience of insiders that expects Ovum to champion the industry (not to be a 'pessimist'). It is a moment of self-conscious performance of membership that draws attention to the fact that market predictions are not neutral. However, the performance of membership does not depend on the self-consciousness: for an indigenous commentator to be pessimistic risks 'talking down' the market.

In this section I have provided an introduction to the AI industry (and its problems) in which I have tried to show how descriptions of an industry may also perform membership of that industry. This applies not only in the notoriously difficult area of market prediction, but even in the identification of the AI industry, of 'knowing' which companies and products were AI. As a former indigenous commentator, I am tempted to say 'but of course that is the AI industry'. In addition, however, the role of the indigenous commentator was to find the words which would represent the situation as resoluble or manageable. In the next section I shall look at ways in which AI companies attempted to manage the problems of the AI marketplace, again exploring ways in which membership of the community is performed through the identification of problems and solutions.

3.2 How AI vendors negotiated the problem of AI

In the previous section I described the beginnings, growth and crisis of the AI industry through the predictive commentary of a series of market reports. In this

section I look at the problem of AI as reported in the specialist AI newsletters (of which the newsletter I edited, *Machine Intelligence News*, was one), which provided a commentary on the analyses and strategies of the AI vendor companies. As journalists, the newsletter editors were not intentionally speaking for the AI companies,[70] but one major source of material was the public announcements and supplementary explanations made by vendors.[71] Reading the newsletters now provides some sense of the relatively unconsidered analysis of a contemporary response to events. The structure of this section is based on the selection of a series of strategic moments in the marketing of AI, as a framework for discussing the ways that AI vendors were represented as attempting to manage the market.

In the summer of 1986 Teknowledge, a Stanford spin-off company set up in 1981 to sell expert-system development environments, announced that it was effectively abandoning the AI programming language Lisp in favour of C,[72] the programming language most used in corporate DP (data processing) and MIS (management information systems) departments. We reported this in *Machine Intelligence News* as the lead story, under the heading 'Teknowledge redescribes the universe' (August 1986, p 1). The significance of this story was that it had been commonly assumed that most expert systems of any interest would be written in AI programming languages (primarily Lisp, but also the European AI language Prolog and some others). Teknowledge had also announced (*MIN* Aug 1986, p 1):

> It is working on a new architecture for AI tools called Copernicus - which no longer views AI as the centre of the universe. The sun in this new cosmology? The IBM world of conventional MIS and DP.

The Teknowledge announcements coincided with the major US trade show for AI, which was combined with the major research conference run by AAAI (American Association of Artificial Intelligence) and called AAAI. Coverage of AAAI 86 in other newsletters had different angles on the same story. *The Spang Robinson*

---

[70] I have in mind here differences between journalism and, say, public relations (PR). This may be seen, for example, in relation to 'ownership' of the text produced from an interview: the normal terms of a journalistic interview include the understanding that the journalist has full control of how the story is handled; in a PR interview, the client expects to set the terms of the story and to approve it before it is published.

[71] For example, I would get information from regular press releases sent out by the vendors, possibly supplemented by a phone interview. For major announcements, vendors would organise press conferences. There were also opportunities for informal discussions at exhibitions and conferences.

[72] Teknowledge had, earlier in the year, announced a C version of its expert system tool S.1, now it was announcing that it would no longer supply the Lisp version (*MIN*, Aug 1986, p 1).

Report had a lead feature called 'IBM validates AI', recording IBM's late commitment to 'the commercial potential of AI', and speculating on the implications for the small specialist AI vendor companies (SRR September 1986). *Applied AI Reporter* headlined its lead story 'AI gets down to business'. All the newsletters reported a speech at AAAI by Herb Schorr of IBM as a sign of IBM's commitment to AI (focusing on expert systems, but talking also of vision, robotics, and natural language processing). Other AI companies had followed Teknowledge's lead to provision of C versions of their tools. Writing in *Applied AI Reporter* (Aug 86, p 3), Tom Schwartz commented:

> The big switch to C continues. Teknowledge is already shipping S.1 in C. Inference is about to show a C version of ART at AAAI. Carnegie Group is going to announce that we can expect C delivery version of Knowledge Craft within two years. Of the gang of four, only IntelliCorp would rather fight than switch ....

The above reports all talk as if AI was about to be taken up in 'mainstream computing'. The switch to C might be taken as an industry getting ready for serious applications. However, with hindsight, it might also be taken as an industry struggling with a recalcitrant market. The same newsletters that reported the entry of IBM into the market also reported problems for the two leading AI hardware companies, Symbolics and Lisp Machines Inc (LMI), two MIT spin-offs selling Lisp workstations and associated software. Symbolics announced that its earnings were up in the first three quarters of the year (fiscal 1986), but it predicted lower earnings for the fourth quarter; LMI was predicting profits by the end of the year, but in the meantime had laid off 60 staff (AAIR Aug 86). Symbolics did indeed report 'marginal' losses in the quarter ended September 30, 1986.[73] In October 1986, LMI lost seven software developers, the larger part of its Process Systems Division who left to form a new company, Gensym. *MIN* commented at the time (Oct 1986, p 1):

> The Process Systems Division is often regarded as one of LMI's main strengths and was responsible for the development of LMI's Picon system, a real-time expert system for uses such as process control ... [74]

By late 1986, press reports in national, non-specialist publications were talking of an

[73] *MIN*, prophetically as it turned out, headlined this story 'Symbolics sneezes'.
[74] Gensym went on to make a successful business selling process control systems based in expert systems and other techniques. In 1999 it is a publicly quoted company.

'AI Winter'.[75] The newsletters, however, tended to either ignore or challenge the idea. For example, *The Spang Robinson Report* in January 1987 (pp 1-4) labelled the AI winter 'a misperception', arguing that it mistakenly assumed that there was a 'homogeneous' AI industry; and that it was based on the performance of just two AI companies, IntelliCorp and Symbolics, which had both suffered 'dramatic quarterly losses ... after consistent profitability'. Following a survey of AI companies, SRR found that:

> All companies regardless of their product niche viewed the current market condition not as a winter or market downturn, but as a market shift.

MIN (Jan 1987, pp 3-4) did a similar survey of the UK AI industry and reported:

> Almost unanimously, the people we spoke to thought that the so-called AI winter is a misperception of what are actually changes in the marketplace - and a misperception based on the performance of a few high-end specialist American companies.

However, LMI went into 'Chapter Eleven' (the first stages of bankruptcy proceedings under US law) at the end of March 1987, and was looking for a purchaser.[76] Symbolics, after three years of increasing revenues and profits (it went public in 1984), began to experience declining revenues and increasing losses from 1987. Teknowledge experienced losses from the final quarter of 1986 onwards.[77] In an industry in which many of the companies were private, not all the losses were immediately revealed. Inference (a private company), for example, did not reveal any losses, but dropped 21 staff in August 1987. By the middle of 1988, another sign of a declining market was the drop in attendance at trade shows. AAAI 88, which had been booked for a large conference centre in St Paul's Minnesota, on the assumption that numbers would keep rising, saw 'an attendance hovering around the 5,000 mark for the third year running' and a smaller exhibition with 92 vendors against 100 the previous year (*MIN* Sep 1988). The vendor companies that had been started by

---

[75] There was, for example, an article in *The San Francisco Examiner* in December 1986 (cited SRR Jan 1987, p 2). In the UK, *The New Scientist* and the computer press ran stories on the AI winter around December 1986 and January 1987 (cf *MIN* Jan 1987, p 3).

[76] It was ultimately bought by a Canadian chip manufacturer called GigaMoss and subsequently disappeared from sight.

[77] Teknowledge's figures for the final quarter of 1986 appeared to show a profit, and this was accepted by most of the newsletters. However, Joan Cochran writing in AIIR (Dec 1986) argued that this was misleading: 'Adjusting for interest income and a change in accounting methods, the first quarter would show a pre-tax operating loss of $743,000 versus a pre-tax profit of $126,000 a year ago.'

academics were increasingly put in the hands of professional management.[78] However, while expert system applications might still be rarer than once expected, they were going into use by the second half of the 1980s. *The Spang Robinson Report* (Aug 1988, p 2), which insisted there was no crisis, only an industry in transition, argued 'AI is a technology not a market': that is, the significance of AI was not as an industry defined in terms of the early pioneering companies, but as a technology (ie an approach to programming and systems design) that was sold by many different companies and incorporated in many different products.

The different responses to the story of the AI winter, by the outsider, national press on the one hand, and the insider, specialist newsletters on the other, might be taken as a moral example of indigenous commentators 'going native' and failing to report the unpleasant news in front of their eyes. However, it is also important to note that the indigenous commentaries did not necessarily hold out comfort for the AI industry or for individual AI vendors. For example, *Spang Robinson's* continuing insistence that there was no crisis, based in the analysis that AI was 'a technology not a market', might be seen as challenging one of the most basic assumptions of the AI industry. Indeed, it could be taken to imply that there should not be an AI industry as such, that the transition required was for the AI industry to disappear, and this was in effect a strategy adopted by some vendors, who began to look for vertical markets. The Carnegie Group, for example, chose to focus solely on the manufacturing market (SRR, Aug 1988, p 6). More generally, the industry attempted to sell AI systems to the corporate marketplace by addressing its installed or 'heritage' computing systems, and indeed the move to C began to seem like the first step in this transition out of the AI marketplace.

The newsletters may have initially discounted stories of the AI winter. However, the unfolding story of crisis in the AI industry was told in the newsletters through their coverage of the manoeuvres of the industry. The final strategic moment to look at in this section concerns a further attempt by AI vendors to capture corporate clients by abandoning AI. In October 1992, *MIN* reported a visit to London by

[78] Russel Noftsker, the MIT academic who had founded Symbolics was replaced as CEO in January 1987. In January 1989, he was unsuccessful in a bid to return to the board simply as a Director, though he continued to hold stock (AIWeek, Jan 1 1989, p 4). When Symbolics eventually went bankrupt in 1998, the assets were bought by a new company called Symbolics Technology headed by Russel Noftsker.

senior executives of Symbolics, 'talking to customers and journalists in what was touted as a "UK relaunch" of the company'. We published a story giving a brief history of Symbolics according to a newly appointed VP of product development. He described the company as abandoning AI in favour of graphics in the late 1980s; then selling its graphics interests to a Japanese company in January 1992;[79] and now (September 1992) turning back to its 'software engineering technology' (once called an 'expert systems development environment') and aiming to sell this as an environment for developing complex real-time applications in heterogeneous environments. The following year, representatives from Inference, IntelliCorp and another leading AI company called Neuron Data visited London within a few weeks of each other, also all announcing corporate relaunches. We ran the stories one after another, to emphasise the similarities, which can be seen in the respective first paragraphs (*MIN*, June 1993, pp 6-7):

> 1)INFERENCE: Senior executives of Inference Corp were in Europe earlier this month, spreading the news of their new marketing philosophy, embodied in ART Enterprise. The new development tool is ultimately based on Inference's well established AI technology, ART; but *on the assumption that AI is only ever going to be a 'niche' market, ART Enterprise is dressed up for the 'corporate applications' market*. This means, essentially, GUIs and objects in *a client-server architecture.*
>
> .....
>
> 2)INTELLICORP: IntelliCorp is also getting its new market strategy in place ... It has launched a new version of Kappa (release 3.0), intended to bring compatibility between Unix and MS-Windows. Applications developed in Unix may now be delivered on PCs. President Ken Haas says, 'Kappa 3.0 is a watershed for IntelliCorp, because *it allows us to take our technology to a much broader audience.*' Kappa 3.0 also provides new GUI tools and C++ interoperability.
>
> At the same time, IntelliCorp is also *targeting client-server architectures* through a new family of products called Kappa CommManager ....
>
> 3)NEURON DATA: Senior management from Neuron Data also recently visited London, to promote the company's new Elements architecture, *aimed at the client-server market*. CEO Patrick Perez believes that the new architecture, with visual programming editors, easier data access across client-server networks, and a new Script language, will open the door to new users, particularly business developers. He says: *'The bottom line here is to broaden our market.'* ... [Emphases NOT in the original]

[79] One Symbolics executive explained the particular vulnerability of graphics system to recession: 'people weren't buying things, which meant the advertisers weren't placing ads, which meant the people who were doing the ads weren't getting the jobs, which meant they weren't buying the systems' (SZ Interview 9/92).

The problem identified by each company was that AI was a niche market, and all three companies shared a new target market: corporate client-server computing. The vendors emphasised that they were not abandoning the AI market (they were, after all, talking to the representative of an AI newsletter). Inference, for example, claimed that its new product (called ART Enterprise) was also the latest release of its AI product ART-IM (in particular, full maintenance ART-IM customers would get a free upgrade to ART Enterprise). At the same time, it was acting to drop its public association with AI. The following month (July 1993, p 3) *MIN* ran a report on AAAI 93 (contributed by David Blanchard of *AI Week*) which ironically remarked:

> Client/server computing literally stole the show at AAAI '93 .... Some of the best known expert system development tool vendors such as Trinzinc (KBMS and ADS), Inference (ART-IM), IntelliCorp (Kappa) and Information Builders (Level5), declined to participate in what used to be - for them, at least - their industry's most important trade show.

The report went on to point out that this was particularly paradoxical, because a conference aimed at industrial developers (Industrial applications of AI, or IAAI), and featuring AI systems built on tools from the AI vendors, was held next door to the AAAI exhibition:

> Inference was boasting that 10 of the 16 applications chosen for the IAAI show were developed using its products, and yet when the corporate attendees of the IAAI show strolled over to the AAAI trade show, Inference was nowhere in sight.

On this evidence, the final strategy of the AI industry was to remove itself from the AI marketplace, and thus to cease to be an AI industry. A web search carried out in 1999 revealed that while many of the companies survive in name, very few acknowledge their history in AI,[80] and very few are in the business of advanced computing.[81] Many of the companies are now offering products in the e-commerce business.[82] With the disappearance of the AI marketplace in the early 1990s, the

---

[80] Some companies, such as Neuron Data and Teknowledge, give a brief history that cites their origins in 'artificial intelligence' and 'expert systems'. Others, however, seem to seek to hide their history. In the case of Inference, I took some time to assure myself this was the same company (not just a purchaser of the name) as the web site includes no history before 1998.

[81] One exception is Carnegie Group, which was acquired by the UK software company Logica. Another important exception is Symbolics Technology, which acquired the assets of Symbolics when it went bankrupt in 1998 and - led by the original CEO and MIT academic Russel Noftsker - is returning to the niche advanced computing market.

[82] For example, Inference now offers software (k-Commerce) to support web-based customer contact centres; Quintus, which was once the major vendor of the AI programming language Prolog, now offers software for corporate call centres (eContact); Neuron Data supplies 'Engines for e-business', rule-based systems for companies to interact with customers over the web; Teknowledge provides a broad range of Internet software and services.

indigenous commentators experienced as many problems as the AI vendors. A few made the transition to a broader, advanced IT arena.[83] Many of the newsletter changed their names in the early 1990s, attempting to locate themselves in a niche market for advanced systems design.[84] Several newsletters went out of business, including *Spang Robinson* and *Machine Intelligence News.*

In summary, this section has looked at the strategic manoeuvres of the AI vendors through the commentary of the AI newsletters, showing how these commentaries first identified the manoeuvres as opening up new opportunities (representing the move from Lisp to C as bringing AI to 'conventional' users), but finally produced a problem of the AI market, for which one resolution was to abandon that market. I have suggested that the newsletter reports of the crisis in the AI industry were specifically distinct from the way that it was reported by 'outsiders' in the national press: stories of an 'AI winter' were reported in the national press, but rejected or qualified in the newsletters; on the other hand, the newsletters stayed with the story, and provide a detailed insider's account of how the unfolding crisis was construed through attempts to manage it by members of the industrial AI community. Earlier (in section 2.3), I compared my situation as a former AI journalist to the situation of indigenous anthropologists as this was presented by Altorki (1982) and others. I suggested that Altorki could challenge western anthropological preconceptions about arab women because she was both 'one of the women' *and* an anthropologist; similarly, I suggested that I could apply an ethnographic analysis to the detailed material that I had gathered as a journalist. It is important to note, however, that the indigenous commentators I have been describing in this section are in a different situation. The newsletters were writing both about and *for* the industrial AI community. Their audience (as well as their subject matter) were a single community. For an indigenous commentator, in this case, configuring the audience may be said to itself involve performing community, as may other rhetorical moves, including the selection of what counts as a story (when and in what way the situation of the market is 'interesting').

---

[83] For example, Ovum achieved this.

[84] For example *AI Week* changed its name to *Intelligent Systems Report*; *Expert Systems Strategies* changed its name to *Intelligent Software Strategies*. *MIN* did not change its name, but like many of the others broadened its remit to cover not only neural networks and areas such as data mining, but also virtual reality systems.

3.3 Explaining the problem of AI

In the two previous sections I have shown how, within the industrial AI community, the problem of AI was adduced as a problem of the market and managed through a series of marketing manoeuvres. In this section I turn to the related but distinct question of how the problem of marketing AI was explained by members of the AI industry, with the purpose of drawing out the implicit theories of technology transfer involved in these explanations. The material in this section is drawn from several subgroups of the industrial AI community and includes interviews with AI vendors, newsletter reports, and also field notes from talks and informal discussion among systems designers at a conference that I attended in 1994.[85] One issue, therefore, will be whether it is possible to see any differences in emphasis among discussions located in different sub-communities.

The explanations for perceived problems of marketing AI tended to fall into two broad but complementary categories - blaming the vendors or blaming the customers. Significantly (in relation to my comments at the end of the previous section about the technological identity of AI as the minimal identity of AI), the technology was seldom blamed. The vendors were usually accused of overselling, and the customers were usually charged with excessive timidity. The following comment - from an interview with a then newly recruited executive at Symbolics (Interview SZ; Sept 1992) - illustrates the accusation of overselling:

> In my opinion (and I've talked to enough people to begin actually to believe it very strongly) AI was oversold. It was oversold by vendors, but mostly it was oversold by the venture capitalist firms and people who were trying to get their return on investment. It was promised to be the answer to all prayers, and clearly it is not that. It does solve certain types of problems. But not every problem can be solved by AI.

By contrast, others blamed perceptions of overselling entirely on the customer, as illustrated by this comment in the newsletter *Intelligent Systems Strategies* (IX.8, Aug 1993 p 2):

> It's unreasonable to blame the early AI vendors for over-promoting their technologies or failing to develop a larger niche for their products. The whole corporate IS culture was stuck with outdated attitudes and techniques, and no one,

[85] The conference, *Advanced Software Solutions in Manufacturing and Engineering*, was held at the Regents Park Hotel in London on 7 December 1994. It was organised by the Prolog Management Group and Compulog Net and was described as a seminar for IT and Technical Managers.

> on the inside or the outside, could change it overnight.

More often, however, the problem of customer perceptions was introduced in the context of a discussion of tips and tactics for managing the customer. The first time I came across the advice not to mention 'AI' was in December 1985, when Karl Wiig of the American consultancy Arthur D Little was a speaker at a BCS expert system conference in the UK. Wiig was describing a Personal Financial Planning System written in Lisp, which created considerable interest in the audience because it seemed to promise the opening of lucrative markets in financial expert systems.[86] *MIN* (Feb 1986, p 6) reported at the time:

> Wiig has only one caution about the various AI techniques used: 'We don't tell the banks it's AI.'

Similar advice is contained in the following story, told to me during conversation at a conference lunch attended mainly by Prolog users and vendors. It was told (Fieldnotes 7/12/94) by a corporate developer with inhouse clients, specifically as a story about the 'problems' that clients have if 'AI' is mentioned:

> In the early days, I suggested an NLP [natural language processing] system for query handling. They said: 'We don't want that here. It's too way out for us'. So I never say how it's done now.

There were two or three other people present, and the story was followed by laughter: the 'problem' (and also the 'joke') was that what the systems developer saw as a relatively straightforward system was regarded by the client as over-ambitious and risky.

There were several variations on these two themes of blaming the vendor (overselling) and blaming the customer (conservatism). One of these was to blame western customers (as opposed to customers in south East Asia and Japan). This may be illustrated by a story in *MIN* (May 1992) about the European launch of a software development tool from Hitachi, called ObjectIQ in Europe, but ESKernel in Japan; it was being marketed in Europe as a tool for object-oriented programming rather than expert systems, on the grounds that European customers were more likely to be put off by the term 'expert system' than Japanese customers.[87] Another

---

[86] In fact another company, Apex, already had an expert financial planning product on the market. In the event, the ADL system never reached the market, nor did the Apex system remain available for long.

[87] Hitachi had commissioned a market survey from Hewlett Packard UK, which advised it on this marketing strategy.

variation is to blame the conservatism of programmers. I heard several examples of this at the conference for Prolog users that I attended in 1994, where the opening speaker (Peter Reitjes of IBM's T J Watson Research Centre) talked of 'the herd instinct' of programmers and their reluctance to explore unfamiliar programming languages. This problem was further explored by one of my neighbours at lunch, a system developer at a Belgian Hospital who said that he uses Prolog for all clinical systems. However, he could not move the main hospital systems to Prolog because 'we have programmers who are, say, 55, and they can only program in cobol. We can't simply sack them' (Fieldnotes 7/12/94). Further variants on blaming the programmer put the blame further back on the conservatism of university teaching practices.[88]

Vendors were prepared to blame the vendors (or at least earlier, other vendors). This may be illustrated by looking further at the explanation given by Symbolics, which I mentioned earlier in this section, in which the company blamed overselling by vendors and, more particularly, venture capitalists, for the problem of marketing AI. For a vendor to be blaming vendors for overselling is not as odd as it might initially seem, particularly when, as in this case, the overselling was placed in the past, and used as a contrast to claims of current customer support (Interview SZ; Sept 1992):

> Historically we were a technology vendor. We used to build technology, the best in the world, and send it all over the world to customers, and hope that they succeeded. Most customers were research. They didn't know how to build the product. Maybe they knew something about computer science, maybe they knew how to do research and build a prototype, maybe they knew something about applications. But they didn't know how to build a large-scale business application and then [grow?] it. Needless to say many of them failed.
>
> The only consultancy ... well, consultancy reported to sales. Consultants would get assigned to do whatever it took to sell the hardware. They built a lot of customised one-off things to get the customer to buy the platform. Once the customer placed the PO [Purchase Order], we shook hands and went our merry way. It is incredibly poor. But at the time when customers were coming to our door, it wasn't necessary to help the customer. We had a huge backlog in 84, 85, 86:

---

[88] One delegate that I spoke to at the Prolog conference (working for a Prolog vendor, but here speaking in the tones of personal enthusiasm), blamed universities for 'reinforcing the hold of sequential programming paradigms'. He said that he had been fortunate to got to a Belgian university where students were 'exposed to ten different programming languages in the first year', but nonetheless he did not appreciate 'the strengths of Prolog' until he came to write his thesis. (Field Notes 7/12/94).

> people were beating a path to Symbolics' door to buy the equipment. We got spoiled. The company was $110 million, thinking it's going to be a one billion dollar company. Not doing anything to get there, but thinking it was going to get there.

This story of customers that did not know what they were doing and a company that did not help them is told to highlight the extensive support and consultancy which Symbolics now provides within the product contract. This is seen, for example, in the observation that Symbolics in the early 1980s gave no advice to customers on how to build 'large scale business applications' - a time when such projects were still advanced research projects. The purpose of the history, however, was to emphasise that Symbolics now (1992) would not think of not giving such support.

A further variation, which perhaps comes close to blaming the technology, but also has implications of overselling concerns rumours of horror stories. As a journalist, I have often been told that there are 'horror stories', attempts to build AI or expert systems that went disastrously wrong, and became expensive mistakes for the client organisation both in money and technological strategy. I have never been told an example of a 'horror story'; and I am not alone in this. At the 1994 meeting (Fieldnotes, 7/12/94), the following interchange took place between my lunchtime neighbours, V (marketing director of a UK AI vendor) and S (systems developer at a large UK manufacturing corporation):

> S: 'And of course there are the horror stories.;
>
> V: '*Are* there horror stories? I'm never really sure about this.'
>
> S: 'Oh yes, I know, it's certainly true.'

Systems designers as customers (ie as users of AI development systems) may have been most inclined to blame the technology. At the 1994 conference, Patric Taillibert of Dassault Electronique presented a paper which described a complex system written in Prolog and combining model-based (AI) and numeric techniques. He contrasted this with an earlier attempt using expert systems techniques which grew to an unmanageable number of rules (1720). The following interchange (in the discussion of Taillibert's paper) illustrates a consensus among the designers of advanced complex systems by the early 1990s (Fieldnotes 7/12/94):

> Q: Of course, expert systems are politically incorrect now, and we've abandoned them. Expert systems got out of control.
>
> PT: There's a problem of maintaining expert systems. When we changed a rule we had to consider all the rules, which was quite impractical. Now [using a new model-

> based approach] we just consider the relevant model. In 1987 we had hoped to use expert systems; three months later we were convinced that expert systems were not adapted to our problem.

It is, incidentally, worth noting that the term 'expert system' is here clearly used to refer to programming techniques; whether expert systems are 'successful' is therefore judged in terms of how well (and perhaps how elegantly) the techniques can be used to carry out a design task. That is, the name 'expert system' is not taken as implying a non-technical description in terms of the simulation of experts (in contrast to the remarks by Ovum (1988b, p 1) quoted in 3.1). In this interchange, 'expert system' names a technical failure.

Finally, I want to turn to another type of explanation, which to some extent redefines the problem. An example of this is given now (1999) by Teknowledge on its web page, within a brief corporate history:

> The Company began producing commercial expert system products in 1984, and went public in 1986. By 1988, Teknowledge competed with over 250 AI companies. The Company learned that while AI is a very powerful technology, it is not an industry in its own right.

This comment (in 1999) that AI is a technology not an industry echoes the comment of the *Spang Robinson Report* in 1988 (Aug) that AI is a technology not a market. It is an odd locution. How could 'a technology' ever come to be mistaken for 'a market' or 'an industry'? Like the accusations of vendors overselling and customer conservatism, this explanation acted to separate 'the technology' from the problem of AI. It reinforced the suggestion that the technology was wrongly marketed, that it should not have been marketed as AI. As vendors and others distanced themselves from the AI market, they distanced themselves from the name; they avoided telling the client 'it's AI'. These moves effectively gave AI a bad name. This phenomenon can be interestingly compared with the power accruing to the name 'Pasteur' which, according to Latour (1988) should itself be regarded as an actant. However, where the name 'Pasteur' apparently had the power to consolidate alliances and gave credence to a variety of proposals and theories, the name 'Artificial Intelligence' seemed to have the opposite effect, and apparently had the power to dispel confidence, undermine proposals and lessen the credibility of a systems designer who was rash enough to mention it to a client.

In the final section of this chapter I discuss whether and how the above explanations of the mis-marketing of AI (which could even perhaps include calling it AI) imply a theory of technology transfer and of communication between social groups.

3.4 Misdescribing and misidentifying: towards a theory of technology transfer?

In the previous section (3.3) I described some of the ways in which the AI community explained the problem of AI and apportioned responsibility. These explanations included blaming the customers, blaming the vendors and, more rarely, blaming the technology. Blaming the customers and blaming the technology may be seen as complementary descriptions, involving a narrative of fearful customers and of vendors that fail to take these fears into account and who make exaggerated claims which they cannot deliver. On the more rarely used explanation of blaming the technology, I suggested first that this was an explanation more likely to be given by a user than a vendor or systems designer (although my example was of a systems designer, speaking *qua* user); secondly, I related the rarity of this explanation to the way in which identifying AI as technology provided a minimal description that the community could fall back on in explaining what AI is. These two points together may be taken to imply a theory of technology transfer as the passing of neutral (technical) information, which may go wrong when technical descriptions are abandoned.

Hype or overselling was only one element of the range of explanations described above for perceived problem of AI. However, it was a common explanation, and it has resonances with charges made by the critics of AI, in the context of the AI debate (which I discuss in more detail in Chapter Six) that researchers' claims for the possibility (or achievement) of machine intelligence were grossly exaggerated (cf Dreyfus, 1979; Putnam 1988b; Searle, 1980; 1984). This suggests a question whether the alleged overselling of AI (by early vendors and others) was in some way related to explanatory practices of the AI research community who described AI in terms of 'intelligence' and through comparisons to human behaviour. It may be tempting to suppose that the early vendors of AI systems did indeed oversell their

products, claiming that they would (or would soon) give human-like performances. This could, further, be blamed on the exaggerations of the academic AI community, who claimed they were building machines that could think like humans. In the chapters that follow, I challenge these suppositions in almost every respect. However, it is also important to see the force of this simple story. It could be used to explain the problems of AI, while allowing vendors to claim that *technologically* there is a neutral description that describes the plain technical potential of their products. The accusation of overselling, placed in the past or laid at the door of other vendors, acts to distance both the overselling and the problems from the product as technological tool. This story also implies a theory of technology transfer as the passing of technological knowledge. One of the ways that technological transfer would go wrong on such a model is the description passed is not accurate (another would be if the recipients fail to understand it). That is to say, the local appropriation of knowledge is only recognised insofar as it is presenting problems. More specifically, as an explanation of what went wrong in the case of AI, it would suggest that the problems arose because people (academics, perhaps, or early vendors) departed from a judicious, technical description of AI, and imported illicit assumptions about the potential of AI, based on unwarranted comparisons with human intelligence. In the following chapters I both test this simple theory, and explore a rather different theory of technology transfer as communication between discursive communities, in which explanations are not easily made to cross discursive boundaries and there may be no commonly available descriptions.

3.5 Summary

In this chapter I have provided a brief introduction to the industrial history of AI. In doing so, I have identified some initial issues in the telling of such a history: I have argued that the narratives told by indigenous commentators may be seen to perform community in adducing their own membership of a community which is at once subject of and audience to the narrative; I have also suggested that community explanations of a perceived problem of AI imply a theory of technology transfer as the passing of neutral, technical information, and that this theory may act as a foil to my own hypothesis that technology transfer involves communication between discursive communities. In 3.1, I described the way that the AI market, its growth

and crisis, were represented through a series of market reports, arguing that the reports performed community in a number of ways but, importantly, through providing an interpretation of the problem of AI that characterised it as solvable or manageable. In 3.2, I described the way that the specialist AI newsletters reported the strategic manoeuvres of AI vendor companies in managing the problem of AI; again, I argued that the newsletters performed community through representing issues of the market in terms that reflected the concerns of the vendors, but terms which were not necessarily less 'objective' than outside commentators. In 3.3, I looked at some of the ways in which AI vendors and systems designers explained the problem of AI, tending to blame either the vendors (for exaggeration) or the customers (for conservatism). These explanations were also reflected in the explanatory practices towards customers (who were not to be told 'it's AI'), and in the marketing strategies of vendors (who ultimately abandoned AI as a marketplace). Through all the above sections, I noticed a tendency to identify AI as a technology or through technical explanations as a means of locating a minimal defensible identity. In 3.4 I suggested that blaming overselling or hype for the problems of AI implied a theory of technology transfer as the passing of neutral or technical descriptions.

In the following chapters of this thesis, I trace the explanatory practices associated with AI through a variety of locations, contexts and configured audiences. One reason for this is implicit in my challenge to the simple story of technology transfer, where the problem is alleged to be overselling based in academic habits of talking about AI as the building of intelligent machines. In the next few chapters, then, I ask just what these habits were, and whether and how AI was explained in terms of the building of intelligent, human-like machines. Another reason, however, has to do with the implications of conceptualising technology transfer as involving communication between discursive communities. As I argued in section 2.4, this implies going beyond a narrow sense of technology transfer (particularly in relation to academic industrial links) and studying the broader context of technology transfer in the history of explanatory practices associated with AI, asking whether and how this history may be reflected in the way that AI was launched on the market. In the following chapter, I begin this study by looking at histories of the early days of AI, asking how the identity of AI as a field was construed for different audiences and in different contexts.

CHAPTER FOUR
NARRATIVES AND AUDIENCE: AI ACQUIRES A HISTORY

4.0 Introduction

At the end of the previous chapter, I suggested a strategy of tracing explanations of AI across different locations, contexts and audiences to produce a history of explanatory practices. The purpose of this strategy is partly to test the theory that AI failed because it was oversold (together with an implicit theory of technology transfer as involving the passing of accurate representations between different groups), but also to provide an understanding of some of the ways in which these explanatory practices differed across locations and contexts, and what this may imply for communication between discursive communities. In this chapter, I look at some historical narratives of the early days of AI, asking how the field was originally identified, and whether this identity may be seen to be constant across different contexts and locations.

In this section, I introduce the chapter by describing the texts that I draw on, and explaining how I divide these into texts for a general audience and those for a peer audience. The differences between these texts is central to my argument, since it provides a way of locating the different narratives and their intended audiences. There is an established tradition of writing books that explain academic fields for a general audience. Here 'audience' may be understood both in terms of the readership categories deployed by book publishers and also in terms of the 'configured audience' as discussed in Chapter Two. Books for a general readership are those books which are not consigned to specific academic fields in the catalogues of academic publishers. The configured audience, however, is that audience successfully addressed within the text. In studying communication between discursive communities, it is the configured audience which is of interest, but the publishers' readership categories help supply the institutional context of publication.

Books explaining ideas of machine intelligence to a general readership have been available at least since the 1950s (eg Weiner, 1950; Sluckin, 1954), although these books described their subject field as Cybernetics. The earliest general books

explaining Artificial Intelligence (under that name) tended to be hostile and written in the vein of exposés, notably Hubert Dreyfus's 1972 book, *What Computers Can't Do*, and Joseph Weizenbaum's 1976 book, *Computer Power and Human Reason.* I return to a discussion of some of the critics of AI in Chapter Six. It is worth noting their early date, however, as they predate the earliest (and perhaps the most famous) of the 'friendly' general books on AI, Pamela McCorduck's 1979 book, *Machines Who Think.* Section 4.1, which looks at the histories told for a general audience, draws on McCorduck's book and a history by Robert Crevier (1993), *AI: The Tumultuous History of the Search for Artificial Intelligence*, as well as some texts by AI academics which address a general readership, such as Herb Simon's (1991) autobiography, *Models of My Life.* Besides being histories for a general audience, these texts (unlike, say, Dreyfus's book) may be described as speaking on behalf of the academic AI community. This is a straightforward point to make in the case of Simon, who was part of the academic AI community. Crevier is a professor of engineering who was on the fringes of AI research.[89] McCorduck, however, is a professional writer. In some respects she may be compared to the indigenous commentators of the industrial AI community described in Chapter Three. However, the very fact that she was writing for a general readership, rather than primarily for the AI community, means that she does not have a service role within the community (which was one of the reasons for describing the indigenous commentators as members of the broad industrial AI community). In addition, she was not a member of a relevant discipline. While one of the claims that I make in this thesis is that it is instructive to view academic disciplines as discursive communities, it may also be noted that they are communities with highly formalised institutional bases. McCorduck is, then, an outsider to the academic AI community. However, like an anthropologist with a favourite tribe, she speaks for the AI community in the sense of telling their story in their terms. She draws on interviews with AI researchers as the main direct source of her history, and many of her chapters are effectively made up of lengthy direct quotations, connected by indirect quotations. Crevier has followed McCorduck's style of research and writing and also follows McCorduck's narrative fairly closely for the period to the late 1970s, and provides very useful

---

[89] Speaking of the MIT AI Lab, which is located in Technology Square, Crevier (1993, pp ix-x) said: 'I spent far too many days and nights in the early 1970s busily assembling a PhD thesis on a topic related to AI, but different enough to keep me away from the lab itself. While doing my own work, I observed the goings on over at Tech Square with a perplexed and somewhat envious interest. ...'

additional material covering the 1980s. There is an additional sense in which McCorduck's history may be said to speak for the AI community, and that is seen in the extent to which her book is cited by members of the community as a historical reference. In one of the odder examples, both Herb Simon (1991) and Ed Feigenbaum (1992) reference McCorduck's description of an interchange between them when Feigenbaum was Simon's student (an interchange for which Feigenbaum was McCorduck's source in the first place).[90] More conventionally, an authoritative textbook, *The Handbook of AI* (Barr and Feigenbaum, 1981, p 5) referenced McCorduck's book as a history, albeit with the qualification that it was 'entertaining'. In the closely related field of Cognitive Science (cf below, 5.1), psychologist Howard Gardner in a 1985 book *The Mind's New Science* (that might be classified either as an introductory textbook or a book for a general audience)[91] explicitly draws on McCorduck's work for a chapter on Artificial Intelligence.

In 4.2, by contrast with the histories for a general audience, I draw on texts addressing a peer or insider audience. These texts are not usually written primarily as histories, but may contain histories. One example is a series of papers written by Marvin Minsky which address a peer audience in the context of research report (1956), a paper presented at a research conference (1959), and a paper published in a reviewed journal (1961). These papers provide a history of AI in the context of an argument for recognising AI as a distinct and important field. Many other texts written for a peer audience include at least a brief description of the field's history (including McCarthy, 1978; Minsky and Papert; 1988; Newell, 1980; 1990; Newell and Simon, 1972; Simon, 1969). Textbooks are also interesting material, since they often address an audience which is not yet fully a peer audience, and indeed part of their purpose is to help enrol students in the discipline. In the case of AI, however, the early textbooks (before about 1980) most addressed a fairly advanced audience (such as graduate students and computer scientists), and the distinction between textbooks and peer texts is not always a clear one in practice. Examples of early AI textbooks include a 1968 collection of papers edited by Minsky (*Semantic Information Processing*), and a 1963 collection edited by Feigenbaum and Feldman

[90] The interchange concerned the announcement by Simon to a graduate class (of which Feigenbaum was a member) in January 1956, 'Over the Christmas break, Allen Newell and I invented a thinking machine' (McCorduck, 1979, p 116; Simon, 1991, p 206; Feigenbaum, 1992, p 194).

[91] There is some overlap between these categories. Indeed, McCorduck's book (which is written from outside the discipline) is frequently to be found in university libraries.

(*Computers and Thought*). I have also drawn on an almost encyclopaedic, multi-authored, three-volume textbook, *The Handbook of AI*,[92] published between 1980 and 1982, which addressed 'colleagues' in industrial research and computer science.

To summarise, for this chapter I have selected two sets of texts, one set providing a history of AI for a general readership, and the other set being texts for a peer audience which include historical remarks. These histories concern the early period of AI, the time when the field was consolidating itself. My assumption is that stories about the early days of any field provide an insight into the identity which is claimed for it. In the case of AI, which as I have previously remarked (2.4) has a strong sense of its own history, the public histories (especially McCorduck, 1979, and the narratives following her history)[93] provide a story which may be seen as a myth of origin. In this Chapter I ask how this myth of origin is construed, and what other alternative historical narratives are available. The more general purpose is to illustrate some of the ways in which AI was identified to different audiences by AI researchers (or those speaking for them) in the early years of AI.

4.1 The first AI conference, the first AI program: History for the general reader

In this section I look at public histories of the origins of AI, that is, histories of AI told for a general reader (as identified in the previous section). I noted in the previous chapter (3.4) that the marketing strategies adopted by AI vendors and systems designers often involved avoiding the use of the name 'AI', and that this strategy acted to endorse 'AI' as a 'bad name', a name that could (to adapt a Latourian phrase) alienate allies. It is interesting to note, therefore, that the public histories tend to emphasise the first use of the name as an effective moment in the constitution of the field of AI. McCorduck (1979) and others (including Crevier

[92] *The Handbook of AI* (Barr and Feigenbaum, 1981; 1982; Cohen and Feigenbaum, 1982) was edited by faculty at the Stanford Department of Computer Science, but included a large cast (a little over 100) of 'chapter editors', 'contributors' and 'reviewers'. While most of these were academics, they included a significant number from industrial research labs, most notably SRI and Xerox, but also individuals from IBM, Fairchild and Honeywell, among others.

[93] I believe that McCorduck is the public source for stories which date AI as beginning in 1956. This date of origin is still sometimes found even in recent documents. For example, in concluding its Intelligent Systems Integration Programme in 1997, the Department of Trade and Industry distributed a CD that provided an overview and history, including the claim (ISIP-CD, p 1) 'The field of computer-based intelligent systems was first defined in its modern form at a conference held at Dartmouth in 1956.'

1993 and Simon 1991), tell the history of the first use of the name 'Artificial Intelligence' as an event which effectively prefigures the field, but can be seen in retrospect to mark its beginning. These stories share a tale about how AI got its name: all agree that the name was chosen or 'invented' by John McCarthy in 1956;[94] they differ only slightly on the 'date of birth' of AI, placing it either in 1955[95] or 1956; and they tend to agree on the names of the main pioneers of the field, and which achievements to include as early breakthroughs, although there are some different emphases. In this section I look at the public history of AI, primarily as told by McCorduck, beginning with descriptions of the choice of name and asking how this contributes to what turns out to be a heroic story of origins.

'Artificial Intelligence' was, McCorduck tells us, first used to name a conference held in 1956. The Dartmouth Conference on Artificial Intelligence is a significant event in McCorduck's history of AI, apparently bringing together for the first time several pioneering AI researchers, 'the first AI program' and the new name. The conference was organised by a group of four people: two young researchers, John McCarthy, then an assistant professor of mathematics at Dartmouth, and Marvin Minsky, a Harvard Junior Fellow in mathematics and neurology; and two industrially based researchers, Nathaniel Rochester from IBM and Claude Shannon of Bell Telephone Laboratories who was already well known for his statistical theory of information and an innovative checkers-playing system.

McCorduck (p 96) tells us that McCarthy chose the name 'Artificial Intelligence', although he may not actually have 'invented' it. She quotes McCarthy's own account:

> 'I won't swear I hadn't seen it before,' he recalls, 'but artificial intelligence wasn't a prominent phrase particularly. Someone may have used it in a paper or a conversation or something like that, but there were many other words that were current at the time. The Dartmouth Conference made it dominate the others.'

The name, he tells us, was available though somewhat marginal: it was the significance of the conference that 'made it dominate the others'. However, there is

[94] Simon's account does not mention McCarthy by name, but credits the group at MIT which McCarthy helped found (Simon 1969, 1981, p 6). McCarthy and Minsky formed the MIT Artificial Intelligence Group after Minsky joined McCarthy at MIT in 1985 (cf Crevier, 1993, p 64). McCarthy was to move to Stanford in 1963 to found an AI lab (Crevier, p 65). In 1963, also, Seymour Papert joined Minsky at MIT, and in 1968 the AI Group became the MIT AI Laboratory (Crevier, p 86).
[95] December 15 1955 according to Simon (1991, p 206).

more to the story. The previous summer, 1955, McCarthy had been working for Shannon at Bell Labs and they had put together a book of collected papers under the title Automata Studies. This name, according to McCorduck (pp 96-97), was Shannon's choice:

> ... McCarthy wanted to use a term different from automata studies for the papers he hoped to get for the book, but Shannon objected that any other phrase was simply too flashy, that the theory of automata would be sober and scientific. McCarthy went along with that, thinking it probably didn't make that much difference.

But McCarthy learned that a name could make a difference:

> ..... Most of the papers they received for the book were in fact about automata theory in the narrowest sense, that is, mathematical principles underlying the operation of electromechanical systems, and not about the relation of language to intelligence, or the ability of machines to play games, or any of the other topics McCarthy was becoming more and more fascinated by.

This anecdote is interesting in its bearing on McCarthy's subsequent choice of the name 'Artificial Intelligence' in preference to 'Automata Theory'. The new name was partly defined by what it was not. Above all, it was not 'Automata Theory', and it was not *like* 'Automata Theory' in several important respects: it was *not* already 'owned', it would *not* attract papers in any established subfield, and it was *not* (the anecdote implies) 'sober and scientific'. These negatives made it available as the name of a conference tackling some advanced and speculative topics, but apparently without a perceived need to present them as 'sober and scientific'.

Shannon was not the only Dartmouth delegate to be unhappy with the name. The problem, however, seems to have been generated by the term 'artificial' rather than implications of non-sobriety. Arthur Samuel, for example, told McCorduck (p 97):

> The word artificial makes you think there's something kind of phony about this, ... or else it sounds like it's all artificial and there's nothing real about this work at all.

Herb Simon made a very similar complaint in an address delivered at MIT in 1968, called 'The Natural and the Artificial' - and now reprinted in a book called *Sciences of the Artificial* (Simon, 1981). As the title of his book indicates, Simon eventually adopted the term 'artificial' with enthusiasm. He had already made the word his own by the time of his 1968 talk, where he uses the distinction natural/artificial to claim a territory for 'the sciences of the artificial' which would cover the behaviour both of humans and of computers. Nonetheless, he registers an objection to the term

'artificial', which is based largely on the dictionary (p 6):

> My dictionary defines 'artificial' as 'Produced by art rather than by nature; not genuine or natural; affected; not pertaining to the essence of the matter'. It proposes, as synonyms: affected, factitious, manufactured, pretended, sham, simulated, spurious, trumped up, unnatural. As antonyms it lists: actual, genuine, honest, natural, real, truthful, unaffected. Our language seems to reflect man's deep distrust of his own products.

In a lengthy footnote (p 6), he first credits researchers at MIT (ie McCarthy and Minsky)[96] for coining the term:

> I shall disclaim responsibility for this particular choice of terms. The phrase 'artificial intelligence', which led me to it, was coined, I think, right on the Charles River, at MIT.

But he does also concede that his own preferred alternatives, such as 'complex information processing' and 'simulation of cognitive processes' may run into dictionary problems of their own:

> for the dictionary also says that 'to simulate' means 'to assume or have the mere appearance or form of, without the reality; imitate; counterfeit; pretend.'

And finally he consoles himself:

> At any rate, 'artificial intelligence' seems to be here to stay, and it may prove easier to cleanse the phrase than to dispense with it. In time it will become sufficiently idiomatic that it will no longer be the target of cheap rhetoric.

(In 1999, this time seems not yet to have arrived.)

On Simon's account, it is the connotations of the word 'artificial' that have made it the target of easy jibes. Ostensibly, the dispute concerns the appropriateness of the name; however it is apparent that the dispute also relates to whose name is being used. In particular, in the late 1950s and early 1960s, 'Artificial Intelligence' is perceived as a name belonging to Minsky and McCarthy, or the MIT group; at Cal Tech, Herb Simon and Allen Newell preferred to call the field 'Complex Information Processing'. The question of whose name it is has echoes of Merton's discussion of eponymy. Merton discussed the role of eponymy in a presidential address made to the American Sociological Society in August 1957, 'Priorities in Scientific Discovery'.[97] He was concerned with the phenomenon of recognition as an

---

[96] Minsky and McCarthy set up the MIT Artificial Intelligence Group in 1958.

[97] This was first published in *American Sociological Review* 22, 6 (December 1957). Page numbers refer to the reprint in Merton (1973).

institutional reward in science, and he identifies 'gradations' of eponymy as one of the major mechanisms of recognising priority of discovery. These gradations, Merton suggests, range from - at the highest - the naming of an age (eg, 'the Newtonian epoch' or 'the Freudian age'), through assignations of the paternity of sciences and disciplines (as in 'Comte, the Father of Sociology'), to 'thousands' of eponymous laws, theories, theorems, etc (as in Boyle's Law). While the name 'artificial intelligence' does not formally include any individual's name, it was, as we have seen, generally perceived as John McCarthy's name or sometimes McCarthy's and Minsky's name.[98] Moreover, rival names also had owners.[99]

If, as Merton suggests, a name may be deployed to reward priority, then a dispute over names might well form part of a priority dispute. And McCorduck does indeed report what looks like a priority dispute involving Herb Simon and Allen Newell on the one side and the organisers of the Dartmouth conference (primarily McCarthy and Minsky) on the other. The dispute broke out in a report-back meeting on the Dartmouth Conference to the Institute of Radio Engineers (at MIT in September 1956), when Simon and Newell (who had attended the two-month conference for just one week)[100] raised objections to John McCarthy (an organiser of the conference) giving the report back.[101] A compromise was reached where McCarthy gave a general report, and Newell and Simon described a program called Logic Theorist that they had presented at Dartmouth. (Simon 1991, 211) Speaking to McCorduck more than twenty years after Dartmouth, both McCarthy and Minsky seem to concede that Simon and Newell had a point. McCorduck reports (p 108):

> McCarthy recalls, 'They felt, perhaps quite correctly, that the situation was anomalous, the conference being reported on by people who hadn't actually done anything, when they had.'

Minsky told McCorduck (p 108):

> The unfairness was that they had a well-developed project that they'd been working

---

[98] Joseph Weizenbaum told Crevier (1993, p 64): 'In the early 1960s, Minsky and McCarthy were almost synonymous, bound together. You never said just Minsky or just McCarthy, you said Minsky-and-McCarthy.

[99] Apart from Simon and Newell's name 'Complex Information Processing', there were also some more remote claims to naming the field; for example, the research group started by Donald Michie in Edinburgh during the 1960s called the field 'Machine Intelligence'.

[100] Simon (1991, p 210) says he and Newell attended the conference for 'about a week'.

[101] The dispute seems to have been acrimonious enough to bring proceedings to a halt for a while. Simon confessed to McCorduck (p 108): '... poor Walter Rosenblith, who was supposed to chair the session, walked around with us on the MIT campus, we strolled down Mem Drive and so on, negotiating.'

> on a long time, pretty much full time, and we'd been working much more casually and much more as generalists for a shorter time, and wanted to share the stage with more or less equal authority which wasn't very nice.

What was it that Simon and Newell had 'done'? Are McCarthy and Minsky making a belated concession of priority, or merely conceding that a working program was an achievement? Simon, in his autobiography *Models of My Life* (1991), describes the development of Logic Theorist in terms which do imply a priority claim. He places the 'birth' of the program six months earlier than the Dartmouth Conference, at the point when he believes they had demonstrated the feasibility of the search technique embodied in Logic Theorist. It is this technique, which he describes as 'heuristic search', which is significant in Simon's account. Indeed it is so significant, apparently, that it deserves to have its birthday remembered (p 206):

> I have always celebrated December 15, 1955 as the birthday of heuristic problem solving by computer, the moment when we knew how to demonstrate that a computer could use heuristic search methods to find solutions to difficult problems.

McCorduck (1979) and Crevier (1993) may be read as giving some support to Simon's claims. Both describe Logic Theorist as 'the first AI program' and suggest that it marked a breakthrough to a 'new paradigm'. McCorduck reports (pp 103-4):

> ...two scientists had arrived on the scene with ... a program embodying the new paradigm, the information processing level of modelling, which would dominate research in artificial intelligence in the next decade.

Crevier, who discusses the Dartmouth Conference in a chapter entitled 'The First AI Program: Defining the Field', agrees. He says (p 48):

> At Dartmouth, Newell and Simon were, as the only participants with a working AI program, far ahead of the others.

However, neither McCorduck or Crevier is really concerned with assigning priority to individuals. Indeed both stress that the importance of the Dartmouth Conference largely lay in bringing people together. McCorduck says (p 109):

> perhaps the most influential result of the Dartmouth Conference itself was the social patterns it set.

Both emphasise the creation of AI as group achievement, and point to the continuing domination of the field by Dartmouth delegates and their students. (McCorduck, pp 109-110; Crevier, p 49) Nonetheless, both deploy the story of the 'first AI program',

present like an unrecognised king at the 'first AI conference'. At the very least, this serves to introduce narrative tension into the story. This is seen if the passage from McCorduck (pp 103-4) is quoted at greater length:

> And that brings us to a fascinating puzzle. For two scientists had arrived on the scene with what no one else had and everyone yearned for - a working and genuinely intelligent program. That alone should have earned them special attention from their colleagues. Perhaps more important, it was a program embodying the new paradigm, the information-processing level of modelling, which would dominate research in artificial intelligence in the next decade. Why wasn't this information-processing level of modelling, as invented by Newell and Simon, recognised at once for what it was?

One of the rhetorical devices used by McCorduck to suggest the destiny of the infant AI is to introduce her knowledge of how things turned out into past tense descriptions of what was happening. In this respect, there is a modal verb which is particularly interesting (pp 103-4):

> a program embodying the new paradigm, the information processing level of modelling, which *would dominate* research in artificial intelligence in the next decade. [my emphasis]

This modality is identified by grammarians as future time in the past (Quirk et al, 1985, pp 218-219).[102] Logically, these constructions do no more than indicate that the narrator speaks with the knowledge of hindsight but, as the grammarians indicate, they thereby often seem to underwrite what they speak of (Quirk et al, p 218):

---

[102] The entry for 'future time in the past' (Quirk et al, 1985, p 219) is worth quoting in full:

'Most of the future constructions just discussed can be used in the past tense to describe something is in the future when seen from a viewpoint in the past.

a) MODAL VERB CONSTRUCTION with *would* <rare; literary narrative style>
The time was not far off when he *would* regret this decision.

b) *BE GOING TO* + INFINITIVE (often with the sense of "unfulfilled intention")
You *were going* to give me your address [... but you didn't...]
The police *were going* to charge her, but at last she persuaded them she was innocent.

c) PAST PROGRESSIVE (arrangement predetermined in the past
I *was meeting* him in Bordeaux the next day.

d) *BE TO* + INFINITIVE <formal>; (i) = "was destined to"; (ii) = "arrangement"
(i) He *was* eventually *to* end up in the bankruptcy court.
(ii) The meeting *was to* be held the following week.

c) *BE ABOUT TO* + INFINITIVE ("on the point of"; often with the sense of "unfulfilled intention")
He *was about to* hit me.

Of all these constructions, only (a) and (di) can be considered genuine expressions of future-in-the-past meaning, in that they alone can be understood to guarantee the fulfilment of the happening in question. For instance:

> Few could have imagined at that time that this brave young officer *was to be*[*/would be*] the first President of the United States of America.

This sentence implies that the young officer (George Washington) did eventually become president of the United States. The other constructions, however (especially (b) and (e)), favour an interpretation of nonfulfilment.'

> [These constructions] can be understood to guarantee the fulfilment of the happening in question.

They are therefore useful constructions both for heightening the sense of destiny and endorsing the veracity of the history told, and in this respect may be seen here as the tools of a somewhat whiggish history (cf 5.3 below). Here, this particular grammatical construction plays its role within a story that foretells glory against the odds. It is, at least, a story that may retain the interest of outsiders.

At the end of this section, then, two more general questions are suggested concerning the performativity of this heroic story of origins. One has to do with the constitution of the reader. Is this a story directed at a general reader only, or does it also appear in insider accounts? To start answering this, I look in the next section (4.2) at some more or less contemporary accounts by Minsky and McCarthy, and other peer commentators, and ask whether they shared the perception that Newell and Simon's program Logic Theorist (LT) 'prefigured the field'. The other question that must be raised concerns the role of a heroic (or other) history in the constitution of the field; this is a question that I return to in Chapter Five.

4.2 'The first AI program' and the scepticism of insiders.

In the previous section, I described how the public histories of AI tend to mark a significant coincidence in 'the first AI program' being presented at 'the first AI conference'. But is this just a tale for outsiders? In this section I ask in what terms, if at all, insiders claim that Logic Theorist prefigures AI as a field? On the face of it, what was important about Logic Theorist was that it was a logic theorem prover,[103] at a time when theorem proving was a recognised challenge.[104] However, for Simon the interest was not in merely proving theorems, but in modelling how humans did it. As he formulated it, the problem was how humans searched for a solution, or what heuristics they used. He explained to Crevier (p 44):

> We were not looking for an efficient way of proving theorems. We were looking at how humans, by selective heuristics, found the right thing to do next.

---

[103] Logic Theorist was eventually able to prove 38 theorems from Russell and Whitehead's Principia Mathematica (Crevier, 1993, p 46)

[104] Minsky was working on a geometry theorem proving system at this time. In the late 1950s a researcher at IBM called Herbert Gelernter developed a Geometry Theorem Prover that could prove theorems involving up to ten steps (Crevier, 1993, pp 55 ff).

By the winter of 1955, Simon, Newell and a third research partner J C Shaw had developed a version of Logic Theorist on index cards, which they played out using the Simon family and several graduate students standing in for program subroutines (Simon 1991, pp 206-207). Implementing Logic Theorist on a computer took another six months, and depended on the development of a programming language called IPL (Information Processing Language) which was based on a technique called "list processing". Simon, in his autobiography (1991, p 212), comments:

> Writing and testing the Logic Theorist was only half of what we had accomplished in 1956. We had also invented a whole new class of computer-programming languages .... These languages were the direct ancestor of John McCarthy's LISP, which has been the standard AI language for thirty years, as well as embodying most of the ideas of what is now called object-oriented programming.

For Simon, then, this was a dual achievement involving both a breakthrough in the modelling of human problem solving and a complementary breakthrough in the design of computer programming languages.

Histories of the field provided by AI researchers for a peer or 'technical' audience usually include mention of Logic Theorist and IPL, and IPL is always mentioned in the history of Lisp. However, the status of 'breakthrough' or 'the first AI program' is not easily conceded. In his 'History of Lisp', written in 1978 for the ACM SIGPLAN History of Programming Languages Conference, John McCarthy says (p 217):

> My desire for an algebraic list processing language for artificial intelligence work on the IBM 704 computer arose in the summer of 1956 during the Dartmouth Summer Research Project on Artificial Intelligence which was the first organised study of AI. During the meeting, Newell, Shaw and Simon described IPL 2, a list processing language for Rand Corporation's JOHNNIAC computer in which they implemented their Logic Theorist program. There was little temptation to copy IPL, because its form was based on a JOHNNIAC loader that happened to be available to them, and because the FORTRAN idea of writing programs algebraically was attractive.

In McCarthy's account, the idea of list processing was one among several inputs that led to the design of Lisp. The idea of an algebraic form of language, as exemplified by IBM's Fortran, seems equally important, and the relative merits of the Johnniac[105]

[105] The Johnniac computer had been designed by John von Neumann for RAND where Simon and Newell began addressing the design of a chess-playing program in 1952. Simon (1991, p 203) remembers that 'with a 4,096-word high-speed store supplemented by a drum with about 10,000 words of usable capacity ..... nothing larger or faster existed.'

and IBM 704 machine architectures are also given a role.

Minsky also tended to downplay the role of IPL and Logic Theorist, both in a series of papers that he produced as a consequence of the Dartmouth conference (Minsky 1956, 1959, 1961)[106] , and in an overview of the field written some ten years later (Minsky 1968). In all these texts, Minsky consistently stressed the continuity between earlier work in cybernetics, games theory, theorem proving and language translation. In the papers following Dartmouth (Minsky 1956, 1959, 1961), he calls the field Artificial Intelligence and describes it (particularly in the first two papers, 1956 and 1959) in terms of 'heuristics' - a term he seems to have learned from Simon and Newell.[107] However, in all three papers, 'problem solving' programs are only one type of heuristic program, although he does describe Logic Theorist (1961, p 21) as

> ... a first attempt to prove theorems in logic, by frankly heuristic methods.

Minsky used the term 'heuristics' in a broadly similar way to Simon and Newell,[108] although in the earlier papers, in particular, he has a rather broad explanation of 'heuristic programming' as the use of 'rules of thumb' in contrast to programming tasks for which algorithmic formalisations are available.[109] He says (1959, p 8):

> 'Hints', 'suggestions', or 'rules of thumb'. which only usually work are called

---

[106] The first of these papers, 'Heuristic Aspects of the Artificial Intelligence Problem' (Minsky, 1956) is an internal Group Report, written soon after the Dartmouth Conference. Simon (1991, p 210) suggests that Minsky's 1956 paper 'reflects very well the general body of knowledge in artificial intelligence that was pooled at the Dartmouth Conference'. The second paper, 'Some methods of artificial intelligence and heuristic programming' (Minsky, 1959), is a version presented to a conference at the National Physical Laboratory in England during November 1958 (NPL, 1959). The final version of the paper, 'Steps Toward Artificial Intelligence' (Minsky, 1961) was published in the peer-reviewed journal Proceedings of the IRE.

[107] Simon (1991, p 199) points out that the mathematician George Polya, who also had an interest in problem solving, employed the term heuristic in a course taken by Newell when he was an undergraduate at Stanford in the 1940s.

[108] *The Handbook of AI* (Barr and Feigenbaum 1981, p 28) suggests that 'the term heuristic has long been a key word in AI'. It distinguishes two different definitions of the term which, it says, 'refer, though vaguely, to two different sets - devices that improve efficiency and devices that are not guaranteed'. It suggests that Newell, Shaw and Simon introduced the definition that opposed 'heuristic' to 'guaranteed' in 1957, and that Minsky (1961) introduced the definition in terms of efficiency. However, while Minsky (1961) certainly emphasises efficiency as a programming issue, earlier versions of the paper (such as Minsky 1959) show that he also raised issues of guarantee in the explication of 'heuristics'.

[109] This explanation of 'heuristic' programming, as a way of getting a computer to undertake a task for which a full formal description is not available, earned Minsky the mockery of the automatic-translation pioneer Yehoshua Bar-Hillel. At the 1958 conference at which Minsky gave this explanation, Bar-Hillel protested against the assumption 'that a machine, if only its programme be specified with a sufficient degree of carelessness, will be able to carry out satisfactorily even rather difficult tasks (NPL, 1959, p 85) (Although Bar-Hillel attributed this belief to McCarthy, McCarthy (NPL, p 90) passed the buck back to Minsky: 'Since other people have proposed this as a device for achieving 'creativity', I can only conclude [Bar-Hillel] had some other paper in mind.')

heuristics. A program which works on such a basis is called a heuristic program.

However, while Simon suggested that heuristic programming was invented (by himself, Newell and Shaw) on 15 December 1955, Minsky provided a list of examples of heuristic programming which included previous work in cybernetics, games theory, theorem proving and so on.

In 1968, Minsky again provided an historical overview of the field of AI, in the Introduction to a volume called *Semantic Information Processing*, which comprised a collection of new work in AI.[110] Again, Minsky located IPL within ongoing work of the 1950s. In this text, he described AI as emerging from cybernetics, and he characterised as 'symbol manipulation' work going on prior to 1956, such as language translation (p 7):

> The serious programming of 'symbol manipulation' processes had started with early experiments on language translation and (somewhat later) on symbolic machine-language compilations; such techniques first became generally available in a form suitable for use by non-specialists in computation with the publication of the IPL programming system in 1956 by Newell, Shaw, and Simon.

Here IPL is represented as a useful tool embodying existing techniques, rather than a breakthrough. Again, the term which he used to characterise the continuing field, 'symbol manipulation' is one associated with Simon and Newell who were developing the concept of Physical Symbol Systems as a unifying model of thinking in humans and computers (Newell and Simon 1972). Minsky then (p 7) went on to make a three-fold distinction in the work which had emerged from the 'cybernetics' of the early 1950s: the first development, he says, related to continuing work in self-organising systems, including simulation of biological mechanisms; but in the second and third approaches that he identifies, Minsky makes an interesting distinction between the Simulation of Human Thought and AI:

> The second important avenue was an attempt to build working models of human behaviour incorporating, or developing as needed, specific psychological theories. .... Work in this area - Simulation of Human Thought - has focused rather sharply at the Carnegie Institute of Technology (Mellon University) where Quillian's work was done, in the group led by Simon and Newell.
>
> The third approach, the one we call *Artificial Intelligence* ,was an attempt to build machines without any prejudice toward making the system simple, biological or humanoid.

[110] Mostly 'slightly edited PhD theses', as Minsky puts it in the Preface to the volume (p v).

A few paragraphs later (p 8), Minsky adds that the content and organisation of the Simulation of Thought and AI are in practice close, partly because

> ... both are in practice using human problem-solving behaviour as their most important model.

This three-fold distinction both indicates a break between the earlier 'cybernetics' work and AI, and it distinguishes AI from more psychological interests. Minsky's history gives Simon and Newell an early and important place in the history of AI. Nonetheless he denied Logic Theorist the title 'first AI program', insofar as it was one heuristic program among others (Minsky 1956, 1959, 1961, 1968) and in addition he distinguished the theoretical concerns that motivated the design of LT as other than 'AI'.

A distinction between AI and the 'simulation of intelligence' was also accepted by Simon and Newell and their pupils. In an overview publication from the Carnegie school (Feigenbaum and Feldman 1963), the contributed chapters are divided into two parts: Artificial Intelligence and Simulation of Cognitive Processes, although Simon and Newell feature in both sections (a paper on chess playing in the AI section and a paper on their problem solving system GPS, the successor to Logic Theorist, in the section on Simulation of Cognitive Processes). In 1972, Newell and Simon published *Human Problem Solving*, a book which provided an influential statement of their theoretical perspective; and this was firmly located within the simulation of intelligence. In the opening paragraph they state (p 1):

> The aim of this book is to advance our understanding of how humans think. ...

The introduction also contains a statement (p 6) of their relation to AI:

> The most important influence upon our choice of tasks such as chess and symbolic logic is the development of the field of artificial intelligence.

They have chosen to focus on tasks that have been studied by AI, they explain, because the AI studies have provided an 'array of plausible mechanisms' for such tasks. This complements Minsky's claim (1968, 8) that the study of human problem solving provides useful models for AI.[111]

Another source of insiders' histories of AI is provided by *The Handbook of AI*. This

---

[111] Paul Cohen in *the Handbook of AI* (Cohen and Feigenbaum 1982, p 7) points out an asymmetry, in that 'AI does not require that an intelligent program demonstrate human intelligence, but information-processing psychologists insist that the correspondence be proved.'

is a three-volume textbook published between 1981 and 1982 by Ed Feigenbaum and other faculty at Stanford, with the aid of a cast of more than 100 named 'chapter editors', 'contributors' and 'reviewers'.[112] (Barr and Feigenbaum 1981, 1982, Cohen and Feigenbaum 1982) Feigenbaum himself was an ex-pupil of Simon's, and the contributors to the textbook were drawn from many academic and industrial centres of AI research.[113] In the *Handbook*, Logic Theorist has a place in a chapter on theorem proving, where it is presented very much as one theorem prover among several (albeit an early one). IPL is described (Barr & Feigenbaum 1982, pp 3-5), in an overview to a chapter largely devoted to the explication of Lisp, as the first programming language to use the list-processing techniques that were later incorporated in Lisp -

> ... the language that has become the mainstay of AI programming.

Although it includes chapters (such as 'Models of Cognition' in Volume Three) which seem to reflect the concerns of the simulation of intelligence, the structure of the *Handbook* does not signal a distinction between AI and the Simulation of Intelligence. This is partly because it is located within AI. The Overview to the chapter on Models of Cognition (Cohen and Feigenbaum 1982, p 3), for example, spends some time explaining that the study of human and machine intelligence can provide mutual insights:

> What we learn about human intelligence suggests extensions to the theory of machine intelligence, and vice versa.

The distinction between AI and the simulation of intelligence is reasserted at several points throughout the *Handbook*. On the other hand, the *Handbook* also reflects a new explanatory practice, which had begun to gain strength in the 1970s, of locating AI as a 'Cognitive Science' which implies links between AI and cognitive psychology, linguistics, and other disciplines. In Chapter Five (5.1) I explore the idea that Cognitive Science constituted an alliance between a number of disciplines, and ask what the implications are for practices of explaining AI.

Insiders' histories of AI tend to steer us away from the easy idea of Logic Theorist

---

[112] Each volume of The Handbook includes a section acknowledging the various researchers who have worked on each section. In some cases this reveals a clear author (eg the Chapter on Cognitive Modelling in Volume Three which may be largely assigned to Paul Cohen), but in others there are many names associated with any one chapter.

[113] Industrial organisations represented included, most notably, SRI and Xerox, but there were also individuals from IBM, Fairchild, and Honeywell, among others. There were almost no researchers from outside the US, one exception being Donald Michie, then of the University of Edinburgh.

as 'the first AI program', a technical breakthrough that appeared in 1956, but was only recognised later for what it was. More interestingly, in tracing these histories, a distinction has been revealed between AI and the simulation of intelligence which researchers on either side invoke, while nonetheless citing each other's work (and contributing to the same journals and conferences). Newell and Simon (1972), like Minsky (1968), described the relationship between AI and the simulation of intelligence as close (for different pragmatic reasons), but nonetheless all three researchers also insisted upon a distinction between the two fields. This has consequences for the proper use of the name: in the peer histories, 'AI' is a name that belongs to one side only, but in the public histories, 'AI' unites both sides. McCorduck and Crevier presented AI as a joint work, an alliance created by the different researchers who first came together in Dartmouth in 1956 (cf section 4.1). The peer histories suggest that the public histories are wrong in their identification of 'the first AI program' and even wrong in their use of the name 'AI'. In the concluding section of this chapter (4.3) I ask how the differences between the public and peer histories should be interpreted.

4.3 Conclusion. Audience as a socio-rhetorical mechanism

In this chapter I have looked at some different ways in which the history of AI has been described in texts for a general audience and in insider texts. For a general audience, the history of AI was presented in a structured narrative, with a date of origin - 1956 - and a fateful bringing together of the name, the people, and the first AI program. In texts for a peer audience (written by some of the very same researchers whose testimonies contributed to the public history), the story is contestable at almost every point: AI did not necessarily begin in 1956; General Theorist was not agreed by everyone to mark the beginning of the symbolic paradigm; nor did all accounts endorse the story that IPL was the direct ancestor of Lisp. Is this because the public histories were simplified in order to make the ideas accessible to a broader audience, while the peer histories showed us the complexity of accurate detail? This appears to explain the significance of the different audiences, and also confirms the idea that successful technology transfer involves the accurate communication of ideas (cf 3.4). Thus, it might be said, it wouldn't be appropriate to trust a text for a general audience in judging the real value of a technology (say AI): the public

histories are, as it were, allowed to dramatise and simplify because they are not 'technical' texts.

However, the differences between the various texts are not all explicable in terms of simplification and accessibility. Is it easier to understand the claim that Logic Theorist marked the introduction of the information-processing paradigm (McCorduck, p 206) than the claim that serious work on symbol manipulation processes began with early experiments on language translation (Minsky, 1968, p 6)? I can see no difference in terms of difficulty, or use of technical concepts. Each is a slightly different version of the origins of symbolic processing, and the main difference seems to be in their roles in the presentation and making of alliances. Minsky represents a continuity between AI in the 1960s and earlier work in language translation. In McCorduck's history, the alliance that is presented reflects the institutional status quo in the late 1970s, when work at such centres as MIT, CMU and Stanford were all represented as 'AI'. However, there is perhaps no easy or single way to characterise the differences between the public and peer histories. For example, the difference between McCorduck's 1979 story of the importance of Logic Theorist and IPL (told through the words of Minsky and McCarthy) conflicts, at least in emphasis, with McCarthy's history of Lisp, written at about the same time (1978). One way of understanding this is through a distinction between stories for outsiders and stories for insiders, where outsiders are told a unifying story about the achievements of a field and insiders continue to engage in factional disputes. Indeed this even explains changes in the reference of the name 'AI', which was used publicly for the field as a whole, while in peer discussions it was used to mark an internal difference (cf above, 4.2). However, on other occasions this distinction can work the other way round. For example, John McCarthy told McCorduck that the name 'Automata Theory' had attracted the wrong sort of papers and that he chose the name 'Artificial Intelligence' for the Dartmouth Conference in order to make a break from Automata Theory. In his funding proposal for the Dartmouth Conference (McCarthy et al, 1955), however, McCarthy cited his paper for the Automata Theory collection as relevant previous work. On this occasion the more public story revealed the factional differences with automata theory, while the story for peer reviewers represented as relevant to the conference any previous work that could be represented as relevant.

The point that I wish to emphasise, in showing the different ways in which the configured audience may be seen as determining the story that is told, is that there are no rules about how the difference works: a general audience may be told a unifying story or a factional one. Nonetheless, the audience makes a difference. Adducing the different audiences is a way of providing an explanation of the different histories. However, the appropriate generalisation is that audience makes a difference - not that audience makes such-and-such a difference. The difference, perhaps, can only be adduced in any specific case (and can always be challenged). The latter point is familiar from discussions of indexicality and interpretive flexibility (cf 2.1). The question that I wish to pursue concerns the presence (perhaps the pervasive presence) of such mechanisms within explanatory practices and within communication between discursive communities. One might call them socio-rhetorical mechanisms, reflecting the way in which rhetorical mechanisms act to produce social relations (examples, following the discussion in Chapter Two, might include configuring the user, performing membership or testing membership). The question that arises is whether the configured audience is the main or primary example of a socio-rhetorical mechanism, or whether others may be found. In the next chapter (Chapter Five) I explore this question in relation to what are sometimes called whig histories - that act to promote a faction and cement or destroy alliances.

CHAPTER FIVE
UNITED AROUND THE SYMBOLIC? ALLIANCES IN THE FIELD

5.0 Introduction

In the previous chapter, I suggested that there were differences in peer and public histories of AI that could be understood in terms of configured audience. However, this does not mean that the different practices of identifying and describing AI can be understood simply in terms of audience. In this chapter I look at some of the ways that AI was identified in the context of relations with other disciplines and the narrative terms through which the disciplines were represented as either 'close' or 'different' (as told both by AI researchers and by members of those other disciplines). In section 5.1 I discuss historical narratives that tell of a broad alliance between a number of disciplines (including Psychology and AI) under the name 'Cognitive Science', and in section 5.2, I look at the rewriting of history from the perspective of research in connectionism. Through both sections I also trace the growing use (through the 1970s and 1980s) of the term 'symbolic' to characterise AI, and ask how this characterisation acted both to build alliances around AI and to question the dominance of the field that was ultimately renamed 'symbolic AI'. Does this perhaps suggest that 'the problem of AI' may be explained through the factionalism that distorted explanations of AI? In the final section, I turn to the more general question of the performativity of history and ask how the various narratives described in this chapter and the previous chapter contribute to a more general understanding of communication between discursive communities.

The material for this chapter is again (as in Chapter Four) largely drawn from historical narratives. The location of these histories is primarily an academic one, including the AI research community, researchers in cognitive science (especially Psychology) and in connectionism. In section 5.1 I begin by drawing on a history of Cognitive Science by Howard Gardner (1985) which may be compared to the public histories of AI that I described in the previous chapter. Like those books, it is both accessible to a general reader and can serve as an introductory textbook. However, the field of Cognitive Science, insofar as it was a field, was disparate and members of each of the constituent disciplines of Cognitive Science were, at least relatively

speaking, outsiders to the other disciplines. As Gardner himself pointed out (p xiii), his book was an early attempt to give a systematic account of the field:

> In the mid 1970s, I began to hear the term cognitive science. As a psychologist interested in cognitive matters, I naturally became curious about the methods and scope of this new science. When I was unable to find anything systematic written on the subject, and inquiries to colleagues left me confused, I decided to probe further.

I ask, first, in what terms Gardner provided such a systematic account, secondly whether the claimed alliance between AI and psychology (and other disciplines) undercut the distinction between AI and the simulation of behaviour (noted in 4.2) and, finally, how narratives of AI as a cognitive science entered into explanations of AI as a field during the 1970s. For exploring the second and third questions, I draw on texts by AI researchers, all primarily addressing a peer audience, including *The Handbook of AI* (Barr and Feigenbaum, 1981; 1982; Cohen and Feigenbaum, 1982) as well as works by individual researchers (Newell and Simon, 1972; Newell, 1980; Minsky 1985).

In section 5.2, I turn to the relationship between AI and the field that was variously called 'cybernetics', 'perceptrons', 'neural nets' or 'connectionism'. I look at narratives of this relationship partly through the historical perspective adopted by several connectionist researchers in the mid-1980s, which is also reflected in the narrative told by a sociologist, Mikel Olazaran (1996). I also draw on texts by researchers in AI (Minsky and Papert, 1988; Papert 1988 and *The Handbook of AI* (Barr and Feigenbaum 1982; Cohen and Feigenbaum, 1982)) and comments by connectionist researchers (eg Clark, 1987; Rumelhart et al, 1986), including an interview of David Rumelhart that I published in my newsletter, *Machine Intelligence News* (April, 1987). Again, most of these texts are written for a peer audience, but a peer audience that includes a cross-disciplinary divide (between AI and connectionism). Olazaran's history is the main exception, in that, in terms of where it is published (the journal *Social Studies of Science*) it is directed at a sociological (science studies) audience, rather than one in AI or connectionism. It is therefore a text which more directly offers a sociological reading of the history of AI.

5.1 AI as a Cognitive Science:

Cognitive Science has its own histories which partly overlap with the histories of AI, and in which Simon and Newell's program Logic Theorist is represented as a breakthrough and a beginning. In this section, I introduce the terms in which the history of cognitive science is usually told, drawing largely on the influential book by Howard Gardner, *The Mind's New Science* (1987). I then ask to what extent the consolidation of Cognitive Science was accepted by AI researchers as providing a location and rationale for their field, and in particular whether this was reflected in the terms under which they described their own field of AI.

Gardner's narrative includes a brief survey of what he calls 'amateur' histories, that is, histories of the field provided by the participants. He comments (p 28):

> Seldom have amateur historians achieved such consensus. There has been nearly unanimous agreement among the surviving principals that cognitive science was officially recognised around 1956. The psychologist George A. Miller ... has even fixed the date, 11 September 1956.

The date mentioned by Miller was the second day of a symposium (the Symposium on Information Theory) held at MIT, at which he heard papers by Simon and Newell (describing Logic Theorist) and the young Noam Chomsky[114] (Gardner, p 28; Miller, 1979). Miller's own symposium paper, given on 10 September 1956 ('The Magical Number Seven ....', on the limits to human short-term memory), is also widely regarded as a seminal paper in the history of Cognitive Psychology. Other witnesses cited by Gardner include three psychologists, Jerome Bruner, Michael Posner and George Mandler. Bruner suggested that the computing metaphor revolutionised Psychology in the 1950s (Gardner, p 29; Bruner 1983). Posner and Mandler both emphasised the influence of ideas of information processing (Gardner, p 29; Posner and Shulman, 1979; Mandler, 1981). Gardner's brief survey concludes (p 29) with a perspective from AI, citing a passage from Newell and Simon's Human Problem Solving (1972), which confirms the significance of psychology and linguistics to the history of their own research.[115]

[114] Chomsky's paper, 'Three Models of Language' (Chomsky 1957), argued against the possibility of a grammar based on Shannon's information theory, and introduced, instead, the notion of linguistic transformations.

[115] Newell and Simon (1972, p 4) also identified the Miller and Chomsky papers as seminal: 'One can date the change roughly from 1956: in psychology by the appearance of Bruner, Goodnow, and Austin's Study of Thinking and George Miller's "The magical number seven"; in linguistics by Noam Chomsky's "Three models of language"; and in computer science by our own paper on the Logic Theory Machine.'

The institutional history of the field was based in cognitive psychology, and in particular at the Harvard Center for Cognitive Studies, which was set up in 1960 by psychologists Jerome Bruner and George Miller. Gardner comments (p 32) that, over the next ten years:

> A list of visitors to the Center reads like a Who's Who in Cognitive Science.

In the mid-1970s, the Sloan Foundation set up a $20 million funding programme in cognitive science. Gardner comments (p 36):

> ...the Sloan Foundation's initiative had a catalytic effect on the field. As more than one person quipped, 'Suddenly I woke up and discovered that I had been a cognitive scientist all my life.' In short order the journal *Cognitive Science* was founded - its first issue appearing in January 1977; and soon thereafter, in 1979, a society of the same name was founded.

Gardner identified six component disciplines of Cognitive Science: Psychology, Linguistics, AI, Anthropology, Neurosciences and Philosophy. However, this was perhaps more a programmatic than a descriptive list, as he had some difficulty in showing enduring 'cognitive' strains in many of these fields. He also made it clear that the major alliance within Cognitive Science was between Psychology and AI.[116]

How does the claimed partnership of AI and Psychology, within cognitive science, relate to the distinction between AI and the simulation of intelligence noted in the previous section? On the one hand, it may seem to bring the two sets of concerns closer. Gardner identifies the influence of computers as one of the 'key features' of cognitive science. He said (p 40):

> While not all cognitive scientists make the computer central to their daily work, nearly all have been strongly influenced by it. The computer serves, in the first place, as an 'existence-proof': if a man-made machine can be said to reason, have goals, revise its behaviour, transform information, and the like, human beings certainly deserve to be characterised in the same way.

The target of this argument is behaviourism which cognitive psychologists were reacting against. (Indeed, the argument is slightly startling until one remembers the behaviourists.) However, cognitive psychologists also had rhetorical mechanisms for distancing themselves from the computer. Another commentator from within AI, Paul Cohen (writing in the Overview to a chapter on Cognitive Modelling in the

[116] He said, for example (Gardner 1987, 136): 'I anticipate the merger of cognitive psychology and artificial intelligence into the central region of a new, unified cognitive science....'

Handbook of AI (Cohen and Feigenbaum 1982, 6)), suggested that:

> For most cognitive psychologists, information processing is a metaphor for human thought, a means of focusing attention on new and interesting questions about the mind. Very few cognitive psychologists have implemented information-processing models - programs - of their theories. Even among those who have, the strong position that the program is itself a theory is not universally accepted.[117]

The claim that the computer is a metaphor for human thought provided cognitive psychologists with a means of distancing their discipline from AI, as did the distinction between 'program' and 'theory'. As Minsky (1968) had claimed a distinction between artificial intelligence and the simulation of human intelligence, so cognitive scientists could claim a distinction between a computer simulation of intelligent behaviour and a theory of how human behaviour works.

Simon and Newell (1972), for their part, explicitly rejected claims that the computer should be thought of as a *metaphor* for human thought. They argued (p 5):

> An information processing theory is not restricted to stating generalities about Man. With a model of an information processing system, it becomes meaningful to try to represent in some detail a particular man at work on a particular task. Such a representation is no metaphor, but a precise symbolic model on the basis of which pertinent specific aspects of the man's problem solving behaviour can be calculated.

Can Simon and Newell, labouring to provide a unifying theory of information processing that would hold for humans and machines, be seen as boundary figures between AI and cognitive psychology? It seems more accurate to observe that both disciplines tended to marginalise them: they were distinguished from AI in terms of their interest in the simulation of human intelligence; but within the categories of cognitive science they were characterised as AI rather than psychology. In the *Handbook*, Cohen (p 7) summed up their situation, which he calls 'information processing psychology':

> Their position is so strong that it defines information-processing psychology almost by exclusion: it is the field that uses methods alien to cognitive psychology to explore questions alien to AI. This is an exaggeration, but it serves to illustrate why there are thousands of cognitive psychologists, and hundreds of AI researchers, and very few information-processing psychologists.

[117] Cohen cited in particular the authors of the program HAM (Human Associative Memory), Anderson and Bower (1973), who stated: 'We make no claim that there is any careful correspondence between the step-by-step information processing in the simulation program and in the psychological theory. ... The claim is sometimes made ... that the program is the theory. This is not the case for HAM, and we wish to make this denial explicit' (Quoted Cohen and Feigenbaum, 1982, p 6).

Cohen went on to suggest that the rise of cognitive science involved a 'weakening' of this position. He seems also to suggest (p 7) that some of the differences between information-processing psychology and cognitive psychology are resolved in cognitive science:

> Recently, the strong position has been relaxed to admit research that does not necessarily prove the correspondence between programs and human behaviour but that has some avowed concern for understanding human behaviour. This research is called *cognitive science* by its practitioners.

But this may be Cohen's own projected programme (or perhaps the voice of the next generation), for Newell and Simon gave no indication of 'relaxing' their position (cf Newell, 1980; 1992).

From Newell and Simon's point of view, Cognitive Science was a home location. They had been promoting their programming ideas among psychologists since the late 1950s. In their own brief history of their field (1972, p 887), they emphasise the significance of a 1958 RAND Summer School for

> ... bringing the work of the Carnegie-RAND group into effective relation with the other main streams of information processing psychology.

As Cognitive Science became an increasingly active forum, Newell argued for the adoption of the Newell-Simon concept of physical symbol systems as a theory of information processing, rather than a mere metaphor. In an address to the first Cognitive Science Conference (organised by the Sloan-funded journal *Cognitive Science* in 1980) he treated the concept of a physical symbol system as one which was both familiar to his audience, yet in need of an introduction (Newell 1980, pp 136-7):

> The concept of a physical symbol system is familiar in some fashion to everyone engaged in Cognitive Science - familiar, yet perhaps not fully appreciated. For one thing, this concept has not followed the usual path of scientific creation, where development occurs entirely within the scientific attempt to understand a given phenomenon. It was not put forward at any point in time as a striking new hypothesis about the mind, to be confirmed or disconfirmed. Rather, it has evolved through a more circuitous root [*sic*]. Its early history lies within the formalisation of logic, where the emphasis was precisely on separating formal aspects from psychological aspects. Its mediate history lies within the development of general purpose digital computers, being thereby embedded in the instrumental, the industrial, the commercial and the artificial - hardly the breeding ground for a

> theory to cover what is most sublime in human thought. The resulting ambivalence no doubt accounts for a widespread proclivity to emphasise the role of the *computer metaphor* rather than a *theory of information processing.*

For Newell and Simon, the Physical Symbol System Hypothesis (PSSH) was an articulated theory, in which symbols played a defined role as the discrete elements of a PSS (connected by a set of relations into a symbol structure) (Newell and Simon 1972, p 20). For many others in AI and cognitive science the idea of symbol manipulation was less theorised (or, as Newell puts it, "not fully appreciated").

Nonetheless, within AI at least, the looser idea of symbol manipulation does seem to have served as an explanatory description which applied across many of the sub-fields that counted as AI in the 1970s.[118] Moreover, one of the effects of the development of the field of Cognitive Science was to help in representations of a broad field of AI, stretching from issues in search, heuristics and programming language design (all of which can be seen to be at least partly descended from the issues surrounding LT and IPL) to areas such as natural language processing and robot vision, and also including the field of cognitive modelling (all included in the three volumes of *The Handbook of AI* (Barr and Feigenbaum 1981, 1982, Cohen and Feigenbaum 1982). For AI, at least (if not the other cognitive sciences), symbolic processing became the explanation which justified the connections between the various components of the field of AI. On the other hand, the ubiquity of the concept of symbol manipulation was not a bar to the continued deployment of a distinction between AI and Psychology, or between programs and theories (cf Cohen and Feigenbaum 1982, 6-7), and for designers of expert systems there remained a choice between modelling humans or simply designing a system that would do the required task (cf Barr and Feigenbaum 1982, 83).

A reading of Minsky's writings from the 1980s suggests that the consolidation of Cognitive Science did bring a renegotiation of the boundary between AI and the modelling of human behaviour which he had argued for in the 1960s (as discussed in section 4.2). For example, one of Minsky's explanations to McCorduck for his failure immediately to recognise the significance of Logic Theorist was that it

[118] Mark Derthick, an AI researcher at the industrially funded research centre, Microelectronics and Computer Technology Corporation (MCC), remarked in 1988 that 'The PSSH is widely regarded as correctly describing the intent of researchers in artificial intelligence ...' (reprinted, Clancey et al, 1994, p 368).

appeared to be psychology rather than AI (McCorduck 1979, 105-6). Minsky also told McCorduck (p 107):

> My interests never again came much to coincide with Newell and Simon - that is, never much at the same time. Again, I did not realise until much later the great joke; at almost every stage, I was in various ways more concerned with human psychology, they with artificial intelligence - but neither of us would have agreed at all with that description.

Minsky grew increasingly more interested in the working of the mind. He did not, however, approach this from the viewpoint either of cognitive psychology or Newell and Simon's information processing psychology, but followed a more idiosyncratic and original path. In 1985, he published *The Society of Mind*, in which he conceptualised the mind as a society of agents (roughly, simple processes), a conceptualisation very different from Newell and Simon's PSSH. This did not alter his perception of AI and psychology as different areas, though he appears to have become more convinced of the practical connections between the two. His definition of Artificial Intelligence in the Glossary to *The Society of Mind* was (p 326):

> The field of research concerned with making machines do things that people consider to require intelligence. There is no clear boundary between psychology and Artificial Intelligence because the brain itself is a kind of machine.

While Minsky may have seen the two fields as converging, however, researchers in either of the fields often viewed this book as somewhat lightweight[119] (cf Clancey et al, 1994, pp 259-333 for a series of reviews by different authors, and Minsky's own reply). Despite Minsky, the distinction between AI and psychology remained available to AI researchers as a way of identifying their field. Crevier (1993, p 249) remarked on the continuing distinction between the two fields, despite the alliance supposedly wrought by cognitive science:

> AI workers and psychologists .... often deeply mistrust each other's work and apply adjectives like 'naive', 'slipshod', and 'irrelevant' to work performed in the other discipline.

If one influence of the field of Cognitive Science was to support the claims of AI to provide a theory of mind, another was to help substantiate the sense of AI as itself a

---

[119] One reviewer, Michael Dyer (Clancey et al, p 259), noted: 'It is unfortunate that cognitive scientists have, for the most part, reacted to Minsky's book as though it were light reading or a minor conversation piece, to be relegated to the coffee table. Minsky himself is partly to blame here, since has has written the book in a style as though it were intended only for those uninitiated to the cognitive sciences.'

unified field. I have suggested that symbol manipulation was often conceptualised in a loose rather than a rigorous sense to explicate the unity of AI. In the latter case, a reservation is still available, allowing members of the community to discriminate between AI and theories of mind. Nonetheless, even the broader or looser conceptualisation of symbol manipulation provided a means of representing the unity of the field.

5.2 From 'AI' to 'symbolic AI': history and power.

In the histories that I have explored so far in this chapter and in Chapter Four, one group of researchers has been identified largely as part of the 'pre-history' of AI, that is researchers in cybernetics. Histories of AI (both peer and public) published in the late 1970s tended to present cybernetics as the pre-history of AI, and to identify a crisis in cybernetics research in 1969, when Minsky and Seymour Papert published their book *Perceptrons*. The Minsky and Papert book was generally taken to have demonstrated the limitations of the 'neural-net' approach to designing intelligent systems. McCorduck (1979, p 89) summarised the impact of *Perceptrons* somewhat triumphalistically:

> And so much for the hope of making a machine think by trying as literally as possible to imitate the brain, the meat machine, at the cellular level.

However, this version of history was itself challenged in the mid 1980s, following a revival of interest in some areas of cybernetics by researchers in neural networks, or connectionism. Crevier (1993, p 103) observes:

> Many neural-network investigators still hold a grudge against Minsky and Papert, whom they blame for delaying the bloom of their discipline until the 1980s.

In this section I ask about the shift of power reflected in the histories of succeeding decades which told, first, of the defeat of the cybernetics or perceptron approach, then of the revival of the cybernetic neural-net approach. My main interest is not in how the histories reflect a shift of power, but in how they relate to a shift of power, in what sense an historical narrative may itself be effective.

One example of how to analyse the force of historical narrative in this case is given by Mikel Olazaran (1996) who employs a distinction (taken from Pinch, 1977) between `research area' and 'official history' modes of description. This enables him

to distinguish the questions of whether Minsky and Papert successfully proved the limitation of neural-net architectures, whether neural-net research work actually came to a halt, and the role of contemporary 'histories'. He answers 'no' to the first two questions and then accuses the 'official history' of AI of starving neural net research of funds (p 641):

> The official history of the debate legitimated the authority structure which was emerging in AI, and was used by the elite of the symbolic approach as a defence strategy against heterodox and 'deviant' interpretations and approaches.

There is a problem in the way in which Olazaran makes the distinction between 'research areas' and 'official history' modes of description. He implies that, although Minsky and Papert were taken to have effected a 'closure' of the debate in 1966, they did not really do so because their 'proof' was later effectively challenged (and, Olazaran further adds, they never claimed as much as some people took them to claim, and the argument was never accepted by some neural net researchers). *Therefore* (he argues) the historical commentaries of the 1970s which endorsed Minsky and Papert's findings were doing something other than disseminating research findings: in effect, he introduces a conspiracy to consolidate the power of 'symbolic AI'. That this is Olazaran's argument may be illustrated in the following paragraph (p 640):

> According to the official history, Minsky and Papert replied to Rosenblatt's overclaiming and showed that progress in neural nets was not possible - and after that this field was largely abandoned. But if, as I have shown here, Minsky and Papert did not quite show that, and if (as I will point out soon) neural nets were not completely abandoned, what was the role of the official history? It is my view that its role can only have been the legitimation of the emergence and institutionalisation of the symbolic approach, which came to be seen as the 'right' approach to AI, and as occupying the whole AI discipline.

Olazaran seems to be reading back the ultimate perceived failure of the Minsky-Papert 'proof' into contemporary discussions of it (a trap for historians of science and technology identified by Kuhn (1962), Bloor (1976) and others). But there is also an assumption written into his account about the power of historical narrative. The power of the 'official history' in this account comes from its use by a powerful group ('the elite of the symbolic approach') and the purpose of its use is to consolidate that power. Instead of saying that there was a consensus that was later challenged, Olazaran takes the success of the later challenge to imply that the

consensus requires further explanation. Nonetheless, there is an important issue which is highlighted by Olazaran's paper, about the power of a history to consolidate the power of a group.

I have already (in the first paragraph of this section) illustrated McCorduck's triumphal history of the 'defeat' of the perceptron model. *The Handbook of AI* adopted a more measured tone, but equally assigned cybernetics, the perceptron and other forms of adaptive or self-organising systems to the past. The editors of the first volume (Barr and Feigenbaum 1981, p 5) commented:

> The ideas of the cyberneticists were part of the Zeitgeist, and in many cases they influenced the early workers in AI directly as their teachers.

And in the third volume (Cohen and Feigenbaum, 1982, pp 378-9) it is reported:

> The original publication of the perceptron model sparked a large amount of research, and a fair amount of speculation, concerning the potential for building intelligent machines from perceptrons. Minsky and Papert (1969) attempted to quiet this speculation by proving several theorems about the limits of perceptron-based learning.

However, perhaps the most telling element of the *Handbook*'s treatment of self-organising systems is the extent to which it failed to describe them at any great length. Moreover, where it did describe such work (for example in Volume Three, pp 325-326), this is represented as not being AI, either from the point of view of AI researchers, or of researchers in adaptive learning:

> In the 1960s, attention moved away from learning towards knowledge-based problem solving and natural language understanding (Minsky 1968). Those people who continued to work with adaptive systems ceased to consider themselves AI researchers; their research branched off to become a subarea of linear systems theory.

A four-page description of perceptrons, a few pages later, was introduced within a section dealing with learning systems in engineering and science, with the comment that 'many of the problems addressed are analogous to those encountered in the design of AI learning systems' (p 382). As Olazaran suggests, this history acts to marginalise work in perceptrons (and other adaptive and self-organising systems). However the marginalisation is largely achieved through the small amount of space devoted to perceptrons, through locating them as not being 'AI' and through the attendant history, including reports of the Minsky-Papert proofs of the limitations

of the approach, and the absence of space given to any suggestion that the Minsky-Papert proofs might be disputable.[120] This illustrates *how* perceptrons were marginalised, including the dissemination of the Minsky-Papert results through factual and authoritative peer reports of those results. That is to say, at the time, a consensus regarded them as marginal.

By the mid-1980s the Minsky-Papert proof was under attack, largely as the result of work coming from the PDP (Parallel Distributed Processing) group at the University of California, San Diego which had been set up in about 1981 (Olazaran, 1996, p 643).[121] The publication in 1986 of *Parallel Distributed Processing* by David Rumelhart and James McClelland, based on work by the UCSD group, was largely taken to have re-established the reputation of neural networks (often under the new name of 'connectionism'). Rumelhart and his colleagues argued against Minsky and Papert that:

> [T]he apparently fatal problem of local minima is irrelevant in a wide variety of learning tasks. [...] In short, we believe that we have answered Minsky and Papert's challenge and have found a learning result sufficiently powerful to demonstrate that their pessimism about learning in multilayer machines was misplaced.[122]

The results of the the PDP group were discussed in academic AI circles and widely accepted as a disproof of the Minsky-Papert case (although Minsky and Papert dispute this in the second edition (1988) of *Perceptrons*).[123] One example from the UK, was the 1987 meeting of the AISB (the Society for the Study of Artificial Intelligence and the Simulation of Behaviour, the British professional association for AI researchers). This opened with a presentation on 'Connectionism and Cognitive Science' by Andy Clark of Sussex University who provided an overview of connectionism for an AI audience and concluded by raising the question of the relation between connectionism and 'conventional AI' (Clark, 1987). David

[120] It is not clear that many arguments critical of the Minsky-Papert line would have been available in the 1960s and 1970s. Most of the critics quoted by Olazaran were interviewed in the late 1980s. He does, however, cite a paper by David Block from 1970, reviewing *Perceptrons* and suggesting limitations on the assumptions made by Minsky and Papert.

[121] In 1979 a conference organised at La Jolla (California) brought together researchers working both in the connectionist and symbolic approaches; the papers were published in 1981, and the PDP group set up at UCSD shortly thereafter (cf Olazaran, 1996, p 643).

[122] Cited Olazaran (1996, p 647).

[123] In the second edition (1988) of *Perceptrons*, Minsky and Papert defended themselves against the accusation that they had prematurely halted work in neural nets, with the claim (p xiii) that this was a necessary pause to consider issues of representation and that 'In any case, the 1970s became the golden age of a new field of research into the representation of knowledge'.

Rumelhart gave an invited talk at this meeting, which is not included in the Proceedings (Hallam and Mellish, 1987). I reported this event and interviewed Rumelhart, giving the readers of MIN (Apr 1987, p 4) a very abbreviated history of connectionism:

> Interest in brain-style processing dwindled in the 60s and 70s, Rumelhart says (after early work by the likes of Oliver Selfridge, F Rosenblatt and Marvin Minsky), partly because AI researchers were able to make interesting contributions to the understanding of cognitive psychology within the limits of computing in a von Neumann environment. But he also blames Minsky and Papert who in their book *Perceptrons* combined a detailed mathematical analysis of perceptrons with 'a very negative rhetoric'. This, he believes, gave the impression that a brain-style approach had been proved to be fruitless.

This paragraph (it seems to me now) captured the revised history so succinctly that it is almost a caricature. That the revival of 'brain style computing' is reported in conjunction with a retort to Minsky and Papert is evidence of the then widespread acceptance of the Minsky-Papert proof. However, this history also assigns AI research to a more limited place - in the above example this place is 'within the limits of computing in a von Neumann environment'.[124] In addition, there was a growing practice of using the name 'symbolic AI' to distinguish the field formerly known as AI from connectionism (also now 'AI'). In other words, the *revised* history is *also* performative, establishing new 'facts' and marginalising previous establishments.

The 'official history' told by AI researchers till the mid-1980s and the revised history told by neural-net researchers and propagated by various commentators (including not only MIN, but also Olazaran) may both be described as whig histories: both act to further the cause for which they speak by presenting it as factual and incontestable. However, I do not mean to identify 'whig histories' as some distinct and reprehensible form of history. I assume that all histories (including the one that I am presenting here) are contestable. The epithet 'whig' points to a tone of realism and certainty; but to the extent that whig histories deny other versions, they are equally susceptible to deconstruction. This is demonstrated by Seymour Papert in his 1988 paper 'One AI or Many?' where he recasts the connectionist version of

---

[124] A 'von Neumann environment' meant a sequential, non-parallel environment. While connectionists often assigned symbolic AI to a von Neumann environment, AI researchers in the early 1980s usually assumed that expert systems and the symbol-processing paradigm would only come into their own in a parallel (or 'non-Von') environment.

history as a fairy story in which two sisters, AI and connectionism, the artificial and natural daughters of cybernetics, compete for the love of the wealthy Lord DARPA (Defense Advanced Research Projects Agency) until the artificial daughter slays the natural one. Papert's target is as much the historical assumptions of AI as those of connectionism, however, and he identifies the underlying problem as the search for a single unifying model of mind (generalisable from simplified models). In this respect, he suggests, the revival of connectionism acts as a vindication of behaviourism against cognitive psychologies and may be seen as a 'return of the repressed'.[125]

Some of the ways in which connectionism, in its turn, came to be taken as the future of computing, may be shown by describing briefly how the revised history of connectionism was disseminated or taken up, not only by AI researchers, but also by the industrial AI community. As I have already suggested, the revised history was presented to UK audiences of AI researchers (via meetings such as AISB) and to industrial AI audiences (via MIN for example) in 1987. The UK activity reflected what was already going on in the US - since this was a US-led revival.[126] Peer discussions in Cognitive Science addressed the relative merits of connectionism and the symbolic paradigm for modelling minds.[127] Industrially, the AI newsletters began reporting neural net developments from early 1988.[128] In November 1988, *AIWeek* (1 Nov, 1988) reported that the first INNS (International Neural Network Society) Conference had attracted 1600 delegates; and that at the conference, Jasper Lupo, Deputy Director of DARPA's Tactical Technology Office announced 'Neural

[125] Papert says (1988, p 9): 'Behaviourism has been beaten down in another version of the Snow White story, but the response of academic psychology to connectionism may turn out to be a classic example of the return of the repressed.'

[126] Academically, work in adaptive systems had been going on outside the US before the setting up of the PDP group - for example, by Igor Alexsander and others at Brunel University. These groups were in some sense outside the battle of the histories. For example, Frank Field, Professor of Cybernetics at Brunel, retained a cyberneticist approach to the field in his 1979 book, *Man the Machine*, where he presented self-organising nets as the paradigm of machine intelligence. He used the term 'Artificial Intelligence', but his book made no reference to the group of researchers who thought they had invented the term in Dartmouth in 1956.

[127] In an overview, in a special edition of the journal, *Cognition* (Vol 28, 1988), Mark Derthick remarked: 'Over the past five years there has been tremendous growth in interest in connectionist theories for their claimed ability to learn automatically from an environment, generalise behaviour to novel situations, gracefully degrade in the face of conflicting input or in the face of internal damage, and their superficial similarity to the organisations and behaviour of massively parallel networks of neurons. These properties are at best less than central in traditional theories of cognitive science, which are based on the manipulation of symbol systems' (Reprinted in Clancey et al, 1994, p 363).

[128] A small neural network industry already existed in the US, the lead company being Hecht -Nielsen Neurocomputers (HNC) which offered its first products, the Anza coprocessor and Anza neurocomputing workstation, in August 1987 (*AIWeek*, 15 Aug 1988, p 4)

networks is a more important technology than the atom bomb'. Lupo also unveiled the results of six-month DARPA Neural Network Study which recommended a $400 million program over eight years; but the funding actually announced, by another DARPA speaker (Barbara Yoon), was a two-year 'seed' funding program (*AIWeek*, 1 Nov 1988, p 1):

> [T]o determine the advantage of neural networks over conventional electronic systems, expert systems and connectionist approaches.

The different technologies were to be compared over three application problem areas: speech recognition, sonar signal recognition and automatic target recognition. In the same issue of *AI Week* (p 8), AI consultant Tom Schwartz, speculating on the strategy behind DARPA's announcements, commented

> [T]his program was the most expedient way to bring peace to the technology factions inside and outside DARPA.

In this judgement of the main funding body, however, it is not only neural nets that has to take its chance as one of many competing technologies for advanced application tasks. Expert systems was also just one of many. The revival of connectionism, along with the consolidation of the idea of Cognitive Science, effected a shattering of explanations of AI. This is illustrated in a 1994 review by M G Dyer of Minsky's (1985) *Society of Mind.* Dyer (who described his own speciality as 'symbolic and connectionist models of language comprehension'[129]) remarked (p 270):

> One can view the society of cognitive science as an 'ecology', in which different founders of AI and their colleagues have 'invaded' and gained control of 'niches'. For example, McCarthy and his colleagues at Stanford, eg Genesereth and Nilsson, hold the 'mind-is-logic' niche; Rumelhart, McClelland and their colleagues hold the 'mind-is-a-connectionist-network' niche; Newell and his colleagues hold the 'mind-is-rule-chunking' niche. Within this ecology, Minsky's book has most affinity to connectionism and conceptual dependency theory.

This passage, it seems to me, stands as a telling obituary to the practice (during the 1970s and early 1980s) of deploying the 'symbolic paradigm' to represent AI as a unified field, and it was the return of connectionism which marked the change.

---

[129] This self-description was made in the contributors' biographies section of the collection in which Dyer's paper was reprinted (Clancey et al, 1994, p 526).

5.3 Conclusion.

In this chapter I have traced the way that AI was identified within the context of the negotiation of alliances, both with friends (in Cognitive Science) and between competitors (AI and connectionism). In section 5.1 I discussed how the consolidation of Cognitive Science from the mid-1970s served as a unifying field which not only united AI and Psychology (and other disciplines) but also helped cement the alliance between factions within AI, and that this alliance had available (as a representation of its unity) the 'physical symbol system hypothesis'. In section 5.2 I discussed how the history of connectionism both revealed a previous anti-connectionist history as being partial and itself acted to rewrite history in favour of itself. The histories that I have discussed in this chapter and the previous one may be summarised as follows:

4.1 A public history of AI that presented a unified field within an unfolding narrative;

4.2 Internal divisions and distinctions that qualified or denied aspects of the public history;

5.1 The history of Cognitive Science in alliance with AI, represented AI as unified around the symbolic paradigm;

5.2 The history of connectionism redefined 'AI' as 'symbolic AI', thereby marginalising the field.

Some of these histories are more whiggish than others. I take a whig history, firstly, to be a history that acts to consolidate the power of the powerful, and to represent what has happened as the most reasonable outcome or the right outcome. This is why Kuhn is often described as attacking whig history through his injunction to return to scientific disputes with an open mind about the outcome.[130] Realist and naturalist voices, which present facts as facts, are the allies of whig history; and this tone of voice may be the most useful way of describing what whig is, since all histories (including this one) are written to convince the reader. The examples I have looked at here illustrate that the rhetoric of whig history may also be deployed in bids for power and in response to attack. The connectionist history, for example, was both a response to what it represented as the unjust marginalisation of

[130] I'd always assumed that Kuhn used the phrase 'whig history' in *The Structure of Scientific Revolutions* (1962). But this does not appear to be the case.

connectionism and a moment in the rehabilitation of connectionism. Similarly, McCorduck's (1979) book, with its heroic tale of a machine (and a science) that was bound to triumph may be seen as a response to the public challenges of authors such as Dreyfus (1972) and Weizenbaum (1976), as well as providing an empowering narrative for AI. Further, not all factional histories are easily described as 'whig' histories; for example, histories which argue one interpretation against another are not always appropriately characterised as whig, perhaps insofar as they acknowledge alternative possible histories (for example, the various attempts to propose or discount the significance of the physical symbol hypothesis, discussed in 5.1, retain implicit references to alternative interpretations). A clue to the best way of understanding why some histories attract the epithet 'whig' may be found in Papert's (1988) description of the connectionist revival as the return of the repressed: whig histories may usefully be characterised as histories which suppress alternative readings (or alternative groups of readers) through a variety of mechanisms, including a realist tone of voice.

It is important to note that there are many other possible narrative threads that I could have pursued in looking at the history of the field of AI. In particular, I could have told more stories of splits and divisions, such as the split between the 'neats' and the 'scruffies' (cf Bundy, 1988, p 89; Crevier 1993, 172-6). McCarthy and his colleagues at Stanford represented the 'neats' and Minsky spoke for the 'scruffies' with his concept of frames (Minsky 1975), which allowed for the opportunistic invoking of properties associated with any frame. Herb Simon, for his part, had no time for either camp.[131] In a footnote in his autobiography, Simon (1991, p 192) commented on McCarthy and his colleagues:

> An influential coterie of contemporary artificial intelligence researchers, including Nils Nilsson, John McCarthy, and others, believe that formal logic provides the appropriate language for AI programs, and that problem solving is a process of proving theorems. They are horribly wrong on both counts ....

Once one starts pursuing splits and divisions, it may seem that any field is always further divisible. But, as we have seen, unifying explanations are also available, for

---

[131] Simon also distanced himself from Minsky's frames-based approach through a claim that there was nothing original in it: 'I have my share of amour propre when it comes to getting credit for scientific discovery ... I've been unable to discover in what respects [Minsky's frames] are an advance over description lists. ... As far as I'm concerned, I've been using frames since 1956' (quoted in Crevier, 1993, p 174).

example in a public history (section 4.1) or within a broader alliance (section (5.1). But by showing, for example, that a unifying history acts to configure the audience as not-a-peer-audience I do not mean to suggest that there are *rules* of deployment, or ways that outsiders *must* be addressed, or *conventions* for carrying on a dispute with one's peers. Sometimes the understanding of why alliances are rejected or claimed may be seen to be unique and specific. For example, in a review of Newell's last major work, *Unified Theories of Cognition* (1990), Minsky (1994a, p 107) attacked the very attempt to provide a unified theory of cognition and argued the need for understanding how

> to build a machine that used several different representations.

Most of the review is based on this argumentative distinction. In the final sentences of the final paragraph, however, Minsky says (pp 107-108):

> From the moment I met him in 1956, Allen Newell was one of my heroes. At least five times his ideas transformed mine; I was never the same person after understanding LT, GPS, HPS, or MERLIN, and now Soar in the context of UTC. This latest work will surely stand as a basic advance in AI's theories of knowledge machines. Reading this monumental book recalled to me my sense of awe in seeing the power of GPS, first in its original form and then in the version with learning. I hope others who read it carefully will have the same experience.

Newell had died before the review was published.[132] Minsky's tribute, which is so different in its claims from any of the remarks he makes about Newell (or Newell and Simon) in the texts that I have quoted so far, cannot be understood merely in the light of obituary conventions. It seems to sum up the whole disputatious relationship between Minsky and Newell, recognising the relationship which the disputation constituted.

Finally, there is one more thread that must be picked out from the explanatory practices discussed in these two chapters. The distinction made by Minsky in his 1968 description of the field, between AI and the simulation of thought, provided a way of identifying AI in contrast to Psychology and in contrast to the simulation of behaviour which was still deployed in descriptions of the field in the *Handbook* in 1980 (cf section 4.2). In discussions of the relationship between connectionism and

[132] The editors comment: 'All of the reviewers of UTC wrote these reviews with the expectation of a response by Newell. Tragically, Newell died just as the reviews were completed.' (Clancey et al, 1994, 5) It is possible that I have read this context into Minsky's final paragraph as the editors also comment elsewhere that 'the reviews themselves appear as originally published' (ibid, xii); however I find it almost impossible not to read Minsky's tribute as a response to Newell's death.

AI in the 1980s, however, the texts I have cited all appear to be concerned with the question of how to model minds (cf section 5.2) and Minsky (1985) reports his own conversion to an interest in modelling minds. Does the idea of AI as an approach to systems design disappear in the 1980s, to be replaced by a consensus that AI is concerned with modelling minds? Turkle (1984, p 251) implies that this is not the case when she remarks that the more speculative (or 'philosophical') ideas of AI were mainly the province of

> a small number of people in AI whose ways of looking at the question of mind and program are starting to have an influence in the world beyond the academy.

She adds in a footnote (p 251):

> The people I discuss in this chapter do not represent the AI community as a whole, but a subset characterised by a sense of the discipline as a theory of mind.

However, the people she refers to include many of the most famous and influential AI researchers. They were active in explaining AI to the world at large, including industrial users and governmental funding bodies. Did these theories enter into explanations of AI in the context of technology transfer? In other words, how did claims and beliefs about the intelligence of AI systems enter into explanations of AI for industrial and government audiences? I approach these questions in the next chapter, first by exploring the charges made against AI researchers of philosophical naivete and exaggeration (6.1), then by exploring examples of the way that 'intelligence' was deployed in explanations of AI and discussion within the AI community about 'intelligence' as an explanatory resource (6.2).

CHAPTER SIX
READING AND INTERPRETATION: ASCRIBING INTELLIGENCE

6.0 Introduction

In the previous two chapters I used the idea of a 'whig history' to suggest that there are competing histories associated with AI that can each be seen to be promoting a factional narrative that endorses one point of view and marginalises others. In particular, the public history of AI served to consolidate the reputation of AI in a unifying narrative that repressed or resolved internal peer disputes; this was emphasised by the eventual rewriting of history from a connectionist perspective that demonised the unifying power of the previous history, which it renamed 'symbolic AI'. My analysis served to show first how explanations of AI differed in relation to audience (Chapter Four), and how they differed in relation to historical context (Chapter Five). In each case I emphasised the power or performativity of the text. In this chapter, however, I reassert the power of the reader. The reader in this case is not the individual reader, but different academic disciplines (or the reader qua member of a discipline), and one of the general points that I explore in this chapter is how, for disciplines, community may be performed and tested through the reading of texts (as well as through historical narratives). I look at this in the context of disputes and debates about whether or when it is appropriate or useful to ascribe 'intelligence' to a computer; this includes the so-called 'AI debate' and some other discussions among AI researchers. In focusing on the question of reading, sociological discussions about interpretive flexibility are relevant. I suggested in Chapter One (1.2; 1.3) that Pinch and Bijker (1989) assume that social interests in an artifact somehow predate or are logically prior to the artifact. One of the issues to be explored in this chapter (particularly in 6.1) is whether, by using the idea of performance of community, it is possible to speak of interests determining an interpretation (reading) without supposing that the interests are prior to texts. The second question to be explored (particularly in 6.2) is whether and how the availability of different interpretations is managed in the context of explanations of AI. For example, does the claim that AI is concerned with building intelligent machines perform community for the discipline (is the belief used as a test of membership) and does this suggest that disciplines are usefully conceptualised as

discursive communities; is it an explanation that is deployed across all contexts and for all audiences; or is its use confined to some sorts of contexts and some sorts of audiences (and if so, how may this be analysed)?

I begin this chapter (6.1) by a selective discussion of issues raised by the AI debate as it took place between AI researchers and philosophers. First, I explore some of the ways in which each side represented the claims of either discipline to have something relevant to say about theories of knowledge and mind. Then I look at differences in readings of a particular text (Turing's description of a test for machine intelligence) and whether this may be described as performing community. Finally I ask what the analyses in this section imply about the relationship of disciplinary interests to interpretations of what counts as an interesting issue. In 6.2 I look at the availability of alternative readings and the argument that interpretive flexibility associated with ascriptions of 'intelligence' may be exploited in explanations of AI. I ask what implications this has in relation to communicating or explaining AI to other groups, and in particular whether different ways of argumentatively exploiting the interpretive flexibility of 'intelligence' may act to configure different readers or audiences. Finally (6.3) I ask how perceived problems of defining 'intelligence' were played out in the explanatory practices associated with AI, and whether the charge of 'exaggeration' made against AI researchers in the context of the AI debate can in practice be abstracted from the performance of discipline. That is to say, is it the case that 'wilful' exaggeration by AI researchers can be dissociated from the performance of discipline; and if so, does it follow that this contributed to the 'problem of overselling AI' identified by some AI vendors (cf Chapter Three).

6.1 Reading and interests: the AI debate

The AI debate, as it came to be known, involved critiques of AI from a number of different disciplinary perspectives, mainly attacking the claim that AI researchers sometimes call 'the machine intelligence hypothesis', that human intelligence can be adequately simulated on a computer. I am not interested in trying to adjudicate this debate, but in seeing how disciplines perform community in the context of the debate. In this section I look at some selected texts associated with the AI debate,

asking to what extent differences of interpretation are associated with different disciplines; for example, how do different disciplines identify what the issues are, what counts as a question, what counts as an appropriate answer, and so on? I focus on the differences between AI and philosophy, as the debate between some AI researchers and some philosophers was one of the most extended episodes of the AI debate; and also because the debate was taken up within philosophy (that is, philosophers are to be found on either side of the question) giving greater opportunity to ask about the terms in which the philosophical issues were identified. I begin by looking at some ways in which members of the two disciplines each represented the appropriate location for solutions of the AI debate; then I study a particular example of different interpretations of what the issues are taken to be in Turing's (1950) description of a test for intelligence.

On the AI side, philosophical interventions in what has come to be known as 'the AI debate' were often seen as an unprovoked attack.[133] In a recent radio discussion, for example, the British neuro-computationist Igor Aleksander, reflecting on some of the themes of the 'bickering' between scientists and philosophers (*Start the Week*, BBC Radio 4,16/3/98), called for:

> ... a healthy form of discussion, rather than saying 'Oh you can't do it at all'.

The AI side has tended to hear its critics as saying, again and again, 'Oh you can't do it at all', and has speculated on causes ranging from a love of mystery and confusion[134] to technical incompetence[135] . Some philosophers have justified a uniquely philosophical interest in the field of AI in terms of a distinction between its 'philosophical implications' on the one hand and the 'technical achievements' of AI researchers on the other. So Dreyfus, for example, in the introduction to *What Computers Can't Do* (1972, p xxxv), begins by making this distinction:

> ... I want to make absolutely clear from the outset that what I am criticising is the implicit and explicit philosophical assumptions of Simon and Minsky and their co-workers, not their technical work.

---

[133] In this and some other respects, the AI debate has some echoes of the more recent 'Science Wars', which involves what are often perceived as an unprovoked attack by some scientists on discussions of science in sociology, cultural studies and philosophy (cf Sokal and Bricmont, 1997).

[134] Igor Aleksander, for example, implies this in his remark (during the radio discussion mentioned above): 'There is a lot of mystique and mystery around discussions that have to do with the brain and the question of consciousness.' (*Start the Week*, BBC Radio 4, 16 March 1998).

[135] cf Minsky that 'the self-made critics of AI have virtually no merit at all ... in technical matters' (Article 14265 in Mail Group comp.ai.philosophy)

He then contrasts the importance of their technical work with the naivete of their philosophical assumptions, pointing to:

> [T]he importance and value of their research on specific techniques such as list structures, and on more general problems such as database organisation and access, compatibility theorems, and so forth. [But ...] their philosophical prejudices and naiveté distort their own evaluation of their results ...

Putnam (1988b)[136] makes a similar distinction. The distinction does several things: it locates the field of AI in 'engineering' and situates AI claims about intelligence (about simulating human intelligence or providing a theory of knowledge) as outside the field; at the same time it implicitly claims theories of knowledge and intelligence as the proper object of Philosophy. Some AI researchers, on the other hand, make a distinction that is in some respects similar to the engineering/philosophy distinction, but acts to validate engineering ('the technical') as the proper location for settling the AI debate. Minsky, for example, in a 1993 contribution to a Mail Group discussion on the AI debate (article 14265 comp.ai.philosophy) says of the critics:

> Anyway, there are plenty of serious 'critics' inside the field, arguing about serious problems of scaling, knowledge structure, etc. Unfortunately, the self-made critics of AI have virtually no merit at all, that is, in technical matters.

The implication is that the question of whether intelligence can be simulated is a technical matter and that philosophical discussion has nothing to offer. McCarthy (1988, pp 305-6) subverts the philosophers' complaints about the philosophical naivete of AI in a comment that assumes that the point at issue is programmability:

> Artificial Intelligence cannot avoid philosophy. If a computer program is to behave intelligently in the real world, it must be provided with some kind of framework into which to fit particular facts it is told or discovers. Here I agree with the philosophers who advocate the study of philosophy and claim that one who purports to ignore it is merely condemning himself to a naive philosophy.
>
> Because it is still far behind the intellectual performance of people who are philosophically naive, AI could probably make do with a naive philosophy for a long time. Unfortunately, it has not been possible to say what a naive philosophy is, and philosophers offer little guidance.

He goes on to complain that it is not even possible to derive ways of representing

---

[136] Putnam (1988b, p 270) says: 'Computer design is a branch of engineering (even when what is designed is software and not hardware), and AI is a subbranch of this branch of engineering. If this is worth saying, it is because AI has become notorious for making exaggerated claims - claims of being a fundamental discipline and even of being an "epistemology". .... AI has so far spun off a good deal that is of real interest to computer science in general, but nothing that sheds any real light on the mind (beyond whatever light may have already been shed by Turing's discussions).'

knowledge from any philosophical theory:

> Either no one in AI (including retreaded philosophers) understands philosophical theories well enough to program a computer in accordance with their tenets, or the philosophers have not even come close to the required precision.

AI researchers, he finally suggests, need to develop their own philosophy, resolving issues of ontology, free will, non-monotonic reasoning and the provision of a realist objectivity to robots. These all sound like topics to which philosophers might contribute (or, indeed, believe they are already contributing to), but, McCarthy complains, philosophers seem to find "that the working systems are too trivial to be of interest" (p 307). By representing what is at issue (what is interesting) in terms of programmability McCarthy thereby marginalises philosophy, which has nothing relevant to say about programmability.

Philosophers also, I assume, may be seen performing disciplinary interests. One example is the identification of philosophical implications in non-philosophical texts. This may be explored through looking at some different ways of reading Alan Turing's (1950) description of a test for deciding whether a computer is to be deemed 'intelligent'. Turing was a mathematician and normally published in mathematics journals.[137] However, the 1950 paper ('Computing Machinery and Intelligence') was published in a philosophy journal *Mind* (then edited by Gilbert Ryle). Turing was therefore publishing outside his own field, although *Mind* had previously published at least one other cybernetics paper (Ashby, 1947)[138] . The paper seems to have arisen in an inter-disciplinary context, in a debate in the Philosophy Department at Manchester University in October 1949, in which Turing had participated (Hodges, 1983, p 415).[139] Its publication may be seen as an attempt to communicate between disciplines (an attempt both by *Mind* in publishing the paper, and Turing in

---

[137] Turing's 1937 paper, 'On Computable Numbers, with an application to the Entscheidungsproblem', which is often credited with responsibility for the birth of modern computing and which conceptualised computability through the Turing Machine, was published in the *Proceedings of the London Mathematical Society.*

[138] The paper was entitled 'The Nervous System as Physical Machine with Special Reference to to the Origin of Adaptive Behaviour'.

[139] Turing had been also involved in informal debate with the philosopher Michael Polanyi and, to a less extent, Karl Popper. At the Manchester meeting, held on 27 October 1949, according to Turing's biographer Andrew Hodges 1983 (pp 414- 415), 'Just about everyone in British academic life with a view to express had assembled. It began with Max Newman and Polanyi arguing about the significance of Gödel's theorem, and ended with Alan [Turing] discussing brain cells with J Z Young, the physiologist of the nervous system. In between, the discussion raged through every other current argument, the philosopher Dorothy Emmett chairing. "The vital difference," she said during a lull, "seems to be that a machine is not conscious".'

addressing an audience primarily of professional philosophers). With this in mind, the opening paragraph is interesting. Turing (p 433) says:

> I propose to consider the question, 'Can machines think?' This should begin with definitions of the terms 'machine' and 'think'. The definition might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another which is closely related to it and is expressed in relatively unambiguous words.[140]

In a journal that regularly published papers devoted to issues of usage, Turing's reference to Gallup polls might be taken to betray some unfamiliarity with the context in which he was publishing.[141] Instead of using a dictionary, or observing usage or consulting his linguistic intuitions, Turing devised a test to stand as an objective definition of 'intelligence'. This is the Turing test, in which an interrogator attempts to distinguish between a computer and a person on the basis of a series of freely chosen questions for a period of five minutes.[142] That is, the Turing test is a functional - or operational - definition, to be agreed beforehand, in order to be able to avoid defining or otherwise explaining the term 'thinking': if a computer can pass this test, let us call it 'intelligent', let us say it is 'thinking'. Read as a functional definition, the point of Turing's paper is a claim about the potential capabilities of computing, and he suggested that by about the year 2000, digital computers would be powerful enough to pass the Turing test.

---

[140] The 'more accurate form of the question', which is not in fact a question, was put by Turing in the following passage (1950, p 442): 'Consider first the more accurate form of the question. I believe that in about fifty years' time it will be possible to programme computers with a storage capacity of about 109, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning.'

[141] The question of how (and why) criteria are 'stretched' in accepting cross-disciplinary papers would make an interesting study, and has implications for a more recent cross-disciplinary publication, the so-called 'Sokal hoax' (cf Sokal, 1996).

[142] The idea of the imitation game is introduced in terms of a game where the interrogator is trying to tell a man from a woman, and where the man is trying to fool the interrogator he is a woman, but the woman is trying to convince the interrogator she is a woman. This has led some commentators to suggest that the point of the Turing Test is for a computer to imitate a woman, although this is based on a single ambiguity and not supported by anything else in the paper (cf Collins, 1990; The Boston Computer Museum Catalogue); Simon Schaffer also apparently supports this view (cf a paper given, with Adam Lowe, at CRICT, Brunel University, 17 February 1999, 'Digital Cabinet of Wonders'). Andrew Hodges, in a web site devoted to Turing (www.wadham.ox.ac.uk/~ahodges/scraptest.html), calls this interpretation 'the Pink Herring'.

It is not entirely self-evident how the Turing test is to be read. Minsky (1959) in one of his early papers attempting to delimit the new field of AI, expressly rejects Turing's use of a functional definition. He argues (p 5) that what counts as 'intelligence' is always changing, so by 'tying' the concept to a particular piece of behaviour we may fall out of step with usage:

> Certainly there are many kinds of performances which if exhibited by a man we would all agree, today, require or manifest intelligence. But would we agree tomorrow? For some purposes we might agree with Turing to regard the same performances in a machine as intelligent. In so doing we would be tying the definition of intelligence to some particular concept of human behaviour.

Turing's biographer Andrew Hodges (1983, p 415), himself a mathematician, describes the test as 'an operational definition of "thinking" or "intelligence" or "consciousness"'. In the philosophical literature, on the other hand, especially in introductory textbooks, the Turing test is used as an illustration of philosophical functionalism, or the thesis that functioning intelligently is equivalent to intelligence. For example, Roger Scruton in his introductory survey, *Modern Philosophy* (1994, p 549) states that

> The original inspiration for functionalism was the 'Turing Machine' described by Alan Turing, 'Computing Machinery and Intelligence, Mind, 1950 [...]. Turing's conjecture is that mental operations are sufficiently like those performed by computational systems to warrant explanation in the same way. In particular, mental questions seem to be iterative, leading to results by the repeated application of algorithmic devices. Maybe the brain is a Turing Machine. Turing proposes a test for artificial intelligence (the Turing test), which is that the machine should be able to match any given human performance: if it can do that, what grounds have we for withholding the description 'intelligent' from machines?

This passage elides two separate papers by Turing, a 1937 paper conceptualising computation in terms of the Turing Machine[143] and his 1950 paper describing the Turing test. This is perhaps not ultimately misleading, if a little careless of historical detail.[144] Unlike the problem that Minsky saw in the test, that it tied intelligence to

[143] The Turing Machine was explicated by Turing (1937) in terms of a scanner reading a tape marked off into squares which might be either blank or contain a 1. The system responded to the scanner reading, according to a table of instructions. The Turing Machine was devised by Turing as a means of conceptualising mechanical computing (in the context of an attempt to resolve a mathematical problem (Hilbert's problem)). It is often described as the beginning of modern computing. It has also often been used as a way explaining the idea of computing to novices (eg Weizenbaum, 1975).

[144] The Turing Machine was perhaps more important in the history of functionalism than the Turing test as Putnam, who is often cited as one of the originators of functionalism (cf Mautner, 1996; Blackburn, 1994; Scruton, 1994) used the Turing Machine analogically in a significant early paper (Putnam, 1959). Putnam later changed his views on functionalism (cf 1988a).

descriptions of particular behaviour, Scruton suggests the machine 'should be able to match any given human performance'. But it is not clear how this would act as an objective test in the sense required of a functional definition. Scruton takes the point of the Turing test to be an illustration of functionalism, rather than merely the production of a functional definition. In another introductory philosophy textbook, Jenny Teichman (1988, p 26) complains that the conditions of the Turing test are arbitrarily overspecified:

> I do not know why Turing said 70 per cent and chose five minutes; nor do I see why he thought these figures indicate that the machine's 'intellectual' behaviour is indistinguishable from the man's. The argument does not seem very convincing, and indeed John Searle has used a somewhat similar thought experiment (the 'Chinese Room') as a way of trying to show that machines don't think.

Teichman, again, takes it that the issue in the Turing test is philosophical functionalism (although she takes Turing to be arguing for functionalism, rather than simply assuming functionalism, while complaining that it is not much of an argument).

The two different readings of the Turing test do not conflict. The point is not that it is wrong to identify Turing as functionalist (or implicitly functionalist), but that the interest of the Turing test is read differently by each discipline. The mathematical or engineering reading takes the test to be a functional definition, and the philosophical reading takes it to be an illustration of functionalism. However, both may agree that it is a functional definition and that it is functionalist; the difference is in assumptions about which is the relevant (or interesting) characterisation. These different readings cannot be put down simply to the taking of sides in the hostilities of the AI debate, since the internal philosophical debate was often waged around the question of functionalism (that is, philosophers on both sides of the AI debate took functionalism to be one of the central issues). Further, as McCarthy indicates in his comment about 're-treaded philosophers' not having much to contribute by way of a programmable philosophy, the irrelevance of philosophy was also a charge by AI researchers even against friendly philosophers.[145] The sociological interest in noticing the difference between the two readings is not to choose between them, in

[145] I also recall a conversation in a pub near Sussex University, between a pro-AI philosopher and an AI researcher, in which the philosopher was describing how Hofstadter's (1979) book *Gödel, Escher, Bach* had first interested him in AI, to which the AI researcher commented dryly that he had found the book 'content-free'.

the sense of deciding that one is right and the other wrong, but to see the way in which disciplinary assumptions are performed through those readings. Following Pinch and Bijker (1989), these readings could be analysed in terms of interpretive flexibility, and the selection of a reading explained in terms of the interests of each discipline. But it should also be noted that such readings serve to negotiate and maintain the identities of the different disciplines. Indeed, being trained in a discipline involves learning how to read appropriately. This may be used to suggest that in the case of academic disciplines (though not perhaps so neatly in the case of other discursive communities), 'interests' may be understood as 'what is taken to be of interest'.

Another important question relates to what this example implies for the possibility of cross-disciplinary communication. Contrasting the two readings shows some of the barriers and defences which each discipline erects to defend its own terms and marginalise the interests of other disciplines. What would it take for AI researchers to see philosophical implications as relevant, or for philosophers to recognise empirical investigations as relevant? John Searle's (1980; 1984) 'Chinese Room' argument seems to suggest that no development in the functionality of computers could have any effect on the philosophical argument, since he starts by supposing that the full simulation of natural language use is possible and that nonetheless such a system would not be said to 'understand' language. Dreyfus (1972; 1992), by contrast, argued against the possibility of simulating intelligence in the case both of chess systems (where he argued that a computer would never match even a good amateur game) and in the case of language use (an argument directed particularly against the possibility of scaling up 'toy' systems, such as Winograd's (1972) SHRDLU system). Winograd effectively accepted Dreyfus' criticisms (cf Winograd and Flores, 1986), while a chess program in 1997 defeated Grand Master and world chess champion Gary Kasparov. Does this mean that Dreyfus was right about language and wrong about chess? (I do not take the answer to either part of this question to be obvious.)

In this section I have illustrated some ways in which AI and philosophy each deploy different assumptions about how theories of knowledge and mind are resolved. The

AI debate provided a context in which AI researchers claimed their field as the appropriate location for settling disputes about intelligence and mind. However, in earlier chapters I have shown that a distinction was maintained, at least in peer discussions, between AI and Psychology. Is this compatible with identifying AI as the location for explaining intelligence? In the following section I explore some of the ways in which 'intelligence' was invoked in explanations of AI and ask whether and how these may be seen as performative of the field.

6.2 Interpretive flexibility: 'intelligence' as an explanatory resource

In the previous section I discussed some different readings of the Turing test, asking whether the differences might be seen to be performative of community. These different readings may also be analysed in terms of interpretive flexibility (cf sections 1.2 and 1.3), but in this case it is not clear that this adds much to the claim that alternative readings are available. It has been suggested, however, that in the case of ascriptions of 'intelligence', alternative interpretations may sometimes be exploited within the deployment of the ascription. Woolgar (1989, p 319) suggests that the AI project of building an intelligent machine sustains itself through playing on the dual possibility of denying or ascribing intelligence to a given machine:

> Instead of bringing research to a close, a 'successful' manifestation of intelligence occasions the redefinition of what, after all, is to count as intelligence.

That is, the mechanisation of 'real intelligence' is always a future task, because the achievement of automating any particular task thereby enables it to be represented as merely mechanical. In contrast to Pinch and Bijker (1989), in this example, closure never comes. There is a moving horizon in which 'intelligence' always sits in the next task to be computerised. This might seem to be an interpretation that was primarily of sociological interest. However, the moving horizon was also noted by the AI community itself, and features in some of the discussions and general explanations of what the field is. The British AI researcher Donald Michie (Michie and Johnston, 1984, pp 17-18), for example, complained about the behaviour of the critics of AI through the following imaginary conversation:

> 'Would it be intelligent if a machine could read a newspaper and give you a summary of its contents?' asks the AI scientist.
>
> 'Certainly!' concedes his critic.

'My student', replies the AI man, 'has just written a program to do that (and it does not cheat simply by printing out the headlines).'
'But how does his program work?' asks the critic with an air of suspicion. After a spell with blackboard and terminal he decides that his suspicion was justified. 'So that's all! I don't call that intelligent.'
There appears to be a feeling that if one understands how something works, it is not intelligent. This leads to the idea coined by Larry Tesler[146] that 'artificial intelligence is whatever hasn't been done yet', placing AI workers in a 'no-win' situation.

Here the moving horizon becomes moving goal posts. The AI scientist, in this dialogue, is trying to establish some prior agreement on what would count as intelligence, to establish a functional definition, but the critic refuses to play by those rules. Minsky, on the other hand, has sometimes attempted to incorporate the moving horizon in explications of 'intelligence'. He does this, for example in an early conference presentation (1959, p 6) where, as I mentioned in the previous section (6.1), he rejected Turing's use of a functional definition, and turned instead to 'usage'. However, Minsky selects just one usage:

> In what situations are we less reluctant to attribute intelligence to machines? Occasionally, a machine will seem to be more resourceful and effective than one might expect from casual inspection of its structure. We may be surprised and impressed and we tend to remain so until through analysis or "explanation" the sense of wonder is removed. ..... In the same way, our judgements of intelligence on the part of other humans are often related to our own analytic inadequacies, and these judgements do shift with changes in understanding.

Several decades later, Minsky (1988, p 307) was still making a similar point:

> For practical purposes we usually tell passers-by this easy definition: 'AI concerns performances that a person needs intelligence to do.' For instance, when Slagle wrote the SAINT program in 1960, that was 'AI', because solving college calculus problems then seemed to need intelligence. However, once Jim Slagle showed us how, such problems somehow no longer seemed to need so much intelligence; in fact it left us wondering why students take so long to learn to solve those kinds of problems.
>
> So, in this sense, the term 'intelligence' itself seems only to describe the moving horizon of our growing understanding of how minds might work.

---

[146] I can recall hearing talk of 'Tesler's Law', that 'AI is whatever has not been done yet' in informal discussion among the industrial AI community. My memory is that (contrary to Michie and Johnson's purpose) the implication of mentioning Tesler's Law was to express reservations about academic AI projects. I have not been able to locate the original reference.

In Minsky's interpretation, human intelligence keeps disappearing, draining out of task after task as we come to understand it.[147] However, in contrast to Woolgar's (1989) observation that the invention of intelligent machines is constantly deferred, Minsky does imply a closure at least in principle: the eventual death of the mystique of 'inexplicable intelligence'. On the other hand, the remark about the students requires some comment. I take it that this remark is a joke (it is difficult not to read it as a joke), but the joke undercuts the idea that the moving horizon changes our understanding of 'intelligence' irrevocably.

If AI researchers, consciously or unconsciously, exploited interpretive flexibility in ascribing intelligence to machines, what did this mean for communication with other groups? Did it enable them to get away with exaggeration (is this the source of the alleged hype and misdescription (cf 3.4))? This may be approached by looking in more detail at some examples of comparing human and machine intelligence in explaining AI. In fact the practice of comparing computers to human intelligence predates AI and is found in the discussion surrounding Babbage's machines (cf Schaffer, 1998), as well as some writings by the pioneers of modern computing (Turing, 1950; von Neumann, 1958) and among early cyberneticists (Wiener, 1950). The funding proposal for the Dartmouth Conference, made famous by the heroic public histories of AI (cf Chapter Four), began with a broad programmatic statement about the possibility of the mechanical simulation of intelligence (McCarthy et al, 1955):

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.

The proposal continues by bringing under the aegis of this project both ongoing work in areas such as 'neuron nets' and areas of interest to the proposers (including issues of programming). The purpose of this document was to convince the reader (ie, the Rockefeller Foundation and other researchers invited to participate) that these various topics were both coherent and interesting as the subject of a summer school. Minsky (1956; 1959; 1961), in a series of three papers (briefly discussed in Chapter Four), provided a more sustained attempt to bring ongoing work under the title

[147] There is evidence of the influence of Minsky's description of 'intelligence' from some widely spaced citations. For example, Armer (1963, p 391) quoted below in this section; and John Seely Brown (1984, p 83) 'One of the founders of Artificial Intelligence once defined intelligence as being that attribute of human behaviour that we admire but do not understand.'

'Artificial Intelligence'. In this context, in a paper (1959) addressing a conference at the National Physical Laboratory in the UK, he introduced the description of 'intelligence' mentioned earlier in this section, where intelligence is used of skills that have not yet been explained. Why does Minsky try to describe intelligence at this point (why not leave it to the audience's general understanding of the term to indicate the sort of tasks he means, as the Dartmouth Proposal had done)? Or, to put the question in a way that can more sensibly be answered, what does this description achieve, in this context? One answer may be found in the figure of the hostile critic, who is addressed soon after Minsky (p 6) has introduced his description of how 'intelligence' is used:

> Many people are hostile to such an investigation, maintaining that creativity (or intelligence) is some kind of 'gift' which simply cannot by understood or mechanised. This view can be maintained only through a constant shifting of definition. As soon as any process or performance has been mechanised, it must be removed ... from the list of creative performances. .... The weakness of the advocate of inexplicable creativity lies in the unsupported conviction that after all machines have been examined some items will still remain on the list.

The hostile person can be portrayed as bent on a fruitless task (although the critic might ask whether Minsky's conviction that the list will be empty is not equally unsupported). In addition, perhaps, the requirement to say what is meant by 'intelligence' may be considered a critical question, since as we have already seen (6.1) there is an assumption on the AI or engineering side that the question cannot be rigorously put, but will be decided by technical developments. Minsky's text may thus be said to include, among its configured readers, a hostile one who must be disarmed. He also pre-empts the situation described by Michie and Johnson, where it is the hostile critic who exploits interpretive flexibility. The NPL conference proceedings[148] suggest that there was some anxiety among some of the audience about claims that intelligence could be fully simulated. This is the response Minsky got from one of the industrial delegates, Dr L C Payne of Decca Radar (p 31):

> I should first like to congratulate the author on what I thought was a most stimulating paper. ... Although there are extremists on both sides I don't think anyone is under any illusion about whether a machine can think in the same sense as a human being can think.

---

[148] The conference proceedings (NPL, 1959) include transcripts of the discussion following each paper. The Preface states (p ii): 'The discussion was recorded and all contributors and authors were asked to edit their contributions. The discussion is reproduced in full.'

Payne went on to suggest that 'the human being can be regarded as a digital computer par excellence' (although he drew the line at supposing computers could handle induction), and his comment about extremists does not seem to be directed at Minsky. Indeed, it reads almost as if he was inserting a disclaimer (on Minsky's behalf as well as his own) to disarm any critics. In his final paragraph, Payne (p 33) observed

> I think a lot of obscure metaphysical thinking surrounds discussions of intelligence today.

This also suggests the lurking figure of the hostile critic. Payne's method of dealing with the hostile critic was to avoid sounding 'extreme'. What Minsky achieved (and Payne failed to achieve) was a programmatic view of the field of AI where no task was ruled out ahead of time, just because some critic said (as Aleksander put it more recently, cf 6.1) 'Oh, you can't do that at all'.

Does the hostile reader have to be pre-empted whenever AI is explained in terms of intelligence? This may be approached through a comparison of publications for a general readership and textbooks addressing a relatively technical readership. In publications for general readers and other public forums such as television, AI was invariably explained in terms of 'intelligence', without any qualification. The most provocative statements of leading AI researchers are more likely to be made, and quoted (reproduced), in general books than in textbooks. The following, from Herb Simon in 1957 (Cited in Crevier, 1993, p 1)[149] illustrates how the general reader might be configured and managed:

> It is not my aim to surprise or shock you - but the simplest way I can summarise is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until - in a visible future - the range of problems they can handle will be coextensive with the range to which the human mind has been applied.

This is the voice of the author as expert. The opening disclaimer is interesting: it prepares the reader to be surprised or shocked, presents what is to come as potentially shocking, but negates or disallows this reading. The reader therefore becomes responsible for her own alarm. The author speaks for science and the reader

---

[149] Crevier gives a year of publication, but no bibliographic details of this quotation, which he uses as the motto to his introductory chapter.

is invited to learn from science.[150]

The authors of AI texts sometimes directly tackled the question of how to address different audiences. For example, Feigenbaum and Feldman's (1963) introduction to AI, *Computers and Thought* attempted to make technical texts directly accessible to a wider readership. It consisted of reprints of papers from technical journals in a number of specialised areas, subdivided into Artificial Intelligence and Simulation of Cognitive Processes,[151] with an introduction for each of the two subdivisions. A final section, "Survey of Approaches and Attitudes" provided an overview of arguments about machine intelligence by Paul Armer, and an overview of the field by Minsky (a reprint of Minsky, 1961). In the Preface to the book (pp vi-vii), readers were given guidance as to the order in which they should read it, depending on whether they were general readers, computer scientists and management scientists, or psychologists and philosophers:

> *For the general reader:* The major introductions to Part I on artificial intelligence and Part 2 on simulation of cognitive processes, the introductory article by Turing [a reprint of Turing 1950], followed by the other articles in a sequence dictated by the tastes of the reader and his competence in the subject matter discussed, and finally the summary and review articles by Armer and Minsky. The Minsky critical review might also usefully be the midpoint in a reading of this collection.
> *For the computer scientist and the management scientis*t: The major introductions, followed by Minsky's critical review. Perhaps of high-lighted interest, Samuel's treatment of learning programs, Tonge's management science application, and the research on theorem-proving programs (Newell, Shaw and Simon, and Gelernter).
> *For the psychologist and the philosopher*: The introduction to Part 2 on simulation of cognitive processes, the articles on problem-solving, verbal learning, two-choice behaviour, concept formation, social behaviour and decision-making, in a sequence dictated by the interests of the reader, and finally the papers on artificial intelligence research.

The general reader, then, might avoid the technical articles altogether.[152] The introduction to the AI section and Armer's overview both argue for machine intelligence. The AI overview (specially written for this volume) presents this as the

[150] A provocative tone may therefore act to control the reader, and simply by ascribing intelligence to machines, the text may be read as provocatively breaking a boundary between machines and humans (cf Mays, 1952; Woolgar, 1989).
[151] The distinction between AI and the Simulation of Cognitive Processes was discussed above (Chapter Four).
[152] It may be noted that Turing's 1950 (Turing test) paper, although grouped among the technical articles, is described as an 'introductory article'.

only scientific position on the subject,[153] and presents the reader with the option of agreeing or being unscientific. However, Armer's overview 'Attitudes toward intelligent machines' (reprinted from a Symposium on Bionics organised by the USAF Air Research and Development Command) addresses the reservations of the 'scientific' reader. This paper is praised elsewhere in the volume for its 'middle of the road approach' (p 387). Its main drift is to counter what Armer calls 'negativist' arguments against machine intelligence, and he expresses concern (p 389) that

> The negative arguments existent today tend to inhibit such research.

He introduces the issues (p 390) with an anecdote that locates negativism among 'builders and users of computers' (his former self included):

> The topic came into prominence in the late 1940s when Babbage's dreams became a reality with the completion of the first large digital computers. When the popular press applied the term 'giant brains' to these machines, computer builders and users, myself included, immediately arose to the defence of the human intellect. We hastened to proclaim that computers did not 'think'; they only did arithmetic quite rapidly.

Anxiety about ascribing intelligence to machines is here blamed, not on the unscientific attitude of the reader, but on the popular press.[154] The reader as scientist is then invited to reconsider the question by viewing intelligence as a continuum, on which people and machines may be distinguished, and on which it may (or may not) turn out there is an upper limit to machine intelligence.[155] This

---

[153] The editors (Feigenbaum and Feldman, 1963, pp 2-3) ask: 'Is it possible for computing machines to think' and answer their own question:
'No - if one defines thinking as an activity peculiarly and exclusively human. Any such behaviour in machines, therefore, would have to be called thinking-like behaviour.
'No - if one postulates that there is something in the essence of thinking which is inscrutable, mysterious, mystical.
'Yes - if one admits that the question is to be answered by experiment and observation, comparing the behaviour of the computer with that of human beings to which the term 'thinking is generally applied.
'We regard the two negative views as unscientifically dogmatic. ....'

[154] Armer repeats this point later in his paper (p 397): 'Exaggerated claims of accomplishments, particularly from the publicity departments of computer manufacturers, have resulted in such a strong reaction within the scientific community that many swing too far in the opposite direction.'

[155] Armer's main concern was to persuade the scientific reader not to pre-judge the limits on automating 'intelligent' tasks, partly in order to enrol members of other disciplines in an interdisciplinary project. He concluded with an appeal (p 405): 'The success of our efforts will depend on how well we do in bringing the various disciplines together and on the number of well-qualified scientists who are attracted to this research area.' In his attempt to enrol other disciplines in the AI project, Armer (pp 399-405) appeals to national interests, including an overview of the state of the debate in Russia, and quotes a Russian professor as saying that Newell and Simon were conservative in suggesting that it would take ten years for a computer to be world chess champion. This foreshadows some of the arguments from national interests put in the late 1970s and early 1980s, except that in the later period the threat was represented as Japan (eg Feigenbaum and McCorduck, 1984), which I discuss below, 7.2.

approach attempts to disarm the reader, not by overruling her as unscientific, but by enrolling her scientific open-mindedness.[156] The paper goes on to claim (p 396) that there is

> [a] mounting list of tasks which can now be carried out on a computer but which we normally consider requiring intelligence when performed by humans.[157]

The idea of the continuum provides a way of evading the problem of providing an operational definition of 'intelligence'. Armer both approves of Minsky's description of intelligence as what has not yet been understood,[158] but chides a critic, Meszar (In a 1953 paper entitled 'Machines Can't Think'), for arguing that any mental process which can be automated is not thinking.[159] As descriptions of usage, Minsky and Meszar seem to be making equivalent claims. The difference is in the way the description is argumentatively deployed, in Minsky's case to keep the field open (the temptation to describe behaviour as 'intelligent' will gradually disappear, as more and more tasks are automated), in Meszar's[160] to attempt to rule it impossible (whatever you automate, you will thereby not have achieved an 'intelligent' machine). Both of these are defensible moves which construe the issue, in the first case, as what can be programmed; in the second case, as how ascriptions of intelligence are deployed. The first case may be seen as an example of the performed interest of AI as a discipline in programmability.

The above discussion suggests that one difference between Woolgar's and Minsky's description of the interpretive flexibility (or moving horizon) of 'intelligence' is the difference between what the general reader is told and what the technical reader is told. The general reader, configured as unscientific and uncritical, will always be unable to challenge the authority of the AI scientist who keeps producing further

---

[156] I use the female pronoun as neutral, according to common practice, though it rings slightly odd for a group ('the users and builders of computers') which was overwhelmingly male (with some notable exceptions, such Admiral Grace Hopper, the designer of Cobol).

[157] The list includes geometry theorem proving, checker and chess playing, assembly line balancing, composing music, designing motors, recognition of manual Morse code, and solving calculus problems (Armer 1963, p 396).

[158] Armer (1963, p 391) says: 'It's easy to underestimate the advances, for "intelligence" is a slippery concept. As Marvin Minsky put it, "You regard an action as intelligent until you understand it. In explaining, you explain away".'

[159] Armer (1963, p 396) quotes Meszar as saying: 'Perhaps the most flexible concept is that any mental process which can be adequately reproduced by automatic systems is not thinking.' Armer comments that this 'gambit' means that the lower bound of what counts as thinking can continually be redefined 'so that it is continually above what machines can do today.'

[160] I have not had access to Meszar's paper, published in Bell Telephone Laboratories Record. However, my point here has to do with how the paper was read by Armer.

stories of future 'intelligent' (and 'conscious' and 'creative') machines. The technically competent reader, on the other hand, is reassured that this is an empirical, programming project which acts merely to demystify unscientific prejudices about 'intelligence'. The hostile critic, meanwhile, seems rarely to be addressed directly. She is present, however, in the pre-emptive arguments of the case for machine intelligence. This is a situation which, I suggest, may be understood in terms of the difference between the intended reader (the reader the author wishes to address) and the configured reader (cf section 2.4): the hostile reader is not intended, however she is configured through the various pre-emptive arguments intended to divert her criticism. Nonetheless, hostile readers may appear where least desired or guarded against. The physicist Roger Penrose (1989, p ix), for example, reported that he was goaded into writing *The Emperor's New Mind* after watching 'proponents of strong AI' on a BBC TV programme. More generally, this is a reminder that readers do not always behave as the text demands.

6.3 Conclusion: Exaggeration as a disciplinary performance?

In this chapter I have looked at explanations of AI in the context of the AI debate. This raised two sets of issues which, initially at least, seem somewhat separate: problems of interdisciplinary communication in the context of an interdisciplinary dispute; and problems raised by attempts to control readings of the claim that machines may be intelligent. In 6.1 I looked at the debate between AI and philosophy as a dispute between discursive communities. I suggested that each discipline negotiated and maintained its own interests through the way in which it adduced issues, posed questions and interpreted and recognised arguments. This revealed some of the ways in which disciplines police their own boundaries, but implied real difficulties for interdisciplinary communication. In 6.2 I looked at some of the occasions on which AI has been explained through claims of the possibility of machine intelligence, and discovered that these explanations may be seen to be exploiting the interpretive flexibility of attributions of 'intelligence' in an attempt to control how this claim was read. This may be understood as an attempt to pre-empt the criticisms of the hostile reader, and is deployed differently for different configured readers (particularly for general readers and for technical readers). This

provides the link between the two sets of issues, since this may be seen as an attempt to impose an appropriate (disciplinary) reading and pre-empt an inappropriate (uninteresting, non-disciplinary) reading.

The discussion in this chapter still leaves open the question of whether practices of explaining AI in terms of intelligence are at the root of 'the problem of overselling' as perceived by some AI vendors (cf Chapter Three). Certainly, in the context of the AI debate, one of the accusations bought against some AI researchers was that they exaggerated their results. It is worth briefly reviewing some of these accusations in the light of discussion in this chapter about performance of community and attempts to manage the reader. For example, Dreyfus (1963, p 46) charges Minsky with leaving out the scare quotes round the word 'understand' in an article in Scientific American, reviewing work in AI. Dreyfus noted that the researcher cited by Minsky (Danny Bobrow) had been

> careful in noting he has given a special meaning to the word 'understands'.

By dropping the quotation marks, Dreyfus charged, Minsky exploits an ambiguity in Bobrow's use of 'understands' as a technically defined term, then goes further and makes claims concerning the 'learning' abilities of the system (p 47):

> Once he has removed the quotation marks from 'understand' and interpreted the quotation marks around 'learning' to mean superhuman learning, Minsky is free to engage in the usual riot of speculation.

Putnam (1988, pp 270-1) made a similar accusation of a wilfully exaggerated claim of the achievement of computer understanding against 'a famous name of AI', who apparently backed down as soon as he was challenged.[161] John Searle in the Reith Lectures (1984, p 30) complained:

> My all-time favourite in the literature of exaggerated claims on behalf of the digital computer is from John McCarthy ...[who] says even 'machines as simple as thermostats can be said to have beliefs'. ... I once asked him: 'What beliefs does your thermostat have?' And he said: 'My thermostat has three beliefs - it's too hot in here, it's too cold in here, and it is just right in here.'

---

[161] Putnam reports (1988, pp 270-1): 'Many years ago I was at a symposium with one of the most "famous names" in AI. The famous name was being duly "modest" about the achievements of AI. He said offhandedly, "We haven't really achieved so much, but I will say that we now have machines that understand children's stories." I remarked, "I know the program you refer to" (it was one of the earliest language recognition programs). "What you didn't mention is that the program has to be revised for each new children's story." (That is, in case the point hasn't been grasped, the "program" was a program for answering questions about a specific children's story, not a program for understanding children's stories in general.) The famous name dropped the whole issue in a hurry.'

Searle invites us to be shocked; but why? (Is it worse to say a thermostat has beliefs than to say a computer has beliefs?) The charge has something in common with Dreyfus' complaint about the removal of the scare quotes: that a conceptual confusion is introduced. McCarthy, however, is not trying to elucidate the concept of 'belief', he is trying to reduce the concept to an operationally useful one. However, are these all cases merely of wilful exaggeration, or is it possible to see them as differences of disciplinary interest? A comment by McCarthy (1979)[162] suggests that they may be both at once:

> Ascribing beliefs to simple thermostats is unnecessary for the study of thermostats, because their operation can be well understood without it. However, their very simplicity makes it clearer what is involved in the ascription, and we maintain (as a provocation to those who regard attribution of beliefs to machines as mere mental sloppiness) that the ascription is legitimate.

A difference in interests or what is interesting (cf 6.1) might be suggested, for example, in the value put on simplification, which is useful for McCarthy but for a philosopher like Searle is a way of missing the point. However, this passage also confirms the impression that McCarthy and some other AI researchers enjoyed provoking their critics.

This raises the question of the relation of a discipline as a discursive community to the slightly more ephemeral groups which inhabit disciplines, and may be characterised as 'provocative' or 'boring', and so on. For example, an outsider may miss a joke that is 'obviously' a joke to insiders - however, while this may be a test of community, it is not usually likely to be a test of disciplinary competence. Moreover, just because McCarthy was being provocative, this does not mean that Searle was wrong to be provoked (being provoked in such circumstances can be a way of reasserting the seriousness of the issue). McCarthy's provocation may be wilful (he didn't have to insist on thermostats being intelligent) but it is not divorced from the performance of discipline as a discursive community - and nor is Searle's 'shock'. A similar point may be made concerning the relation between the performance of discipline and claims about machine intelligence. Putnam (1988), as I mentioned (above 6.1) distinguished between the engineering achievements and

---

[162] I accessed this paper in its web, rather than its paper version. The page reference for the web version is page 1 of www-formal.stanford.edu/jmc/ascribing/node4.html.

philosophical claims of AI, arguing that AI had done little that was of any philosophical interest. In the same paper he went on to say (p 277):

> [A]rtificial Intelligence as we know it doesn't really try to simulate intelligence at all. Simulating intelligence is only its notional activity; its real activity is writing clever programs for a variety of tasks.

What is captured by the phrase 'notional activity' is the way in which the simulation of intelligence is available to be invoked as the aim of AI, but that it is also possible to distance the discipline from that aim by explaining technical aims in technical terms. Because the 'technical' explanation is only available for use with a very specific (ie technically trained) audience, the explanation in terms of the simulation of intelligence was used by AI researchers, in lieu of being able to to give a rigorous (technical) explanation. In this sense, it might be said, it was always a boundary explanation, carrying with it assumptions about the priority of the technical, which configured the reader either as an insider who is complicit in the compromise involved in a non-technical explanation, or an outsider who is either nervous or hostile, or thrilled at the fantastic future. The danger of this strategy was that the reader would hear the claims as outrageous (as, from some perspectives, they were). In the context of a discussion of technology transfer, however, the general reader and the philosophical reader are less important than an audience of industrial strategists and policy makers. Were these audiences being sold the intelligent machine or just some very clever programming? I address this in the following chapter.

CHAPTER SEVEN
INDUSTRIAL AND POLITICAL STRATEGIES: AI AS STRATEGIC TECHNOLOGY

7.0 Introduction

My assumption in this thesis (cf 2.4) has been that communication between discursive communities involves both a history of representational practices and a multiplicity of socio-rhetorical purposes and achievements. In chapters Four to Six I have explored some of the explanatory practices deployed by AI researchers and those speaking on their behalf, looking at differences between the way AI was explained to general readers and to technical readers (4.1; 4.2), and how attempts were made to build alliances with other disciplines, and to forestall rivals and critics (5.1; 5.2). In addition, I have shown how attempts were made to manage the perceived 'slipperiness' of the concept of 'intelligence', and in doing so to control the reader (6.2). I discuss some of the general theoretical and analytical issues raised by this study in Chapter Eight. My purpose within the case study, however, is to consider the communication involved in technology transfer narrowly understood (cf 2.4), that is: how was AI explained to, and understood by, industrial users and policy-makers? In this respect, Chapters Four to Six serve a dual contextual purpose: on the one hand to provide a broader context of comparison or contrast, to help in the analytic perception of formulations specifically deployed in the narrower sense of technology transfer; on the other to locate those formulations within a history of explanatory practices. In addition, Chapters Four to Six also serve to provide a fuller picture of the way that social relations are performed through communication between discursive communities. However, my aim is not to present the history of explanatory practices as a chain of influences, like a biblical genealogy. My intention is to show how different sorts of explanatory practices were deployed in different specific contexts and used to co-produce, say, audience, community, facts, authority, reassurance (and so on, in any number of combinations). The generalisation that I wish to make is that the socio-rhetorical aspects of explanation are not 'rhetorical' flourishes, or 'cultural' weaknesses which could and should be dropped in the interests of accurate communication. I return to this in Chapter Eight, arguing that the model of technology transfer as the passing of

neutral or discrete items of knowledge is flawed in its understanding of the social practices of communication and sets up an inappropriate ideal of communication.

In the preceding chapters, I have (in historical terms) gone some way beyond the point when AI was first brought to market, that is, in the early 1980s. One reason for this, again, is to give a fuller picture of the variation in explanations of AI, within a changing historical context. However, it may be useful briefly to locate the first attempts at exploiting AI in the early 1980s in relation to some of the events and discussions covered in earlier chapters. In particular, in terms of the discussion in Chapter Five, it may be noted that in 1980 the 'symbolic paradigm' was at its height, and connectionism, in the guise of perceptrons, was widely considered a dead end. In terms of publications: McCorduck's evangelising history had been published in 1979; and for technical audiences *The Handbook of AI* (Barr and Feigenbaum, 1981; 1982; Cohen and Feigenbaum, 1982), was in process of being completed. My aim in this chapter is to look at texts describing AI to industrial and political audiences, asking how it was explained and understood. I begin (7.1) by looking at some of the ways in which AI was explained to industrial researchers, configured as technical colleagues, particularly through *The Handbook of AI.* I then ask how these discussions may be compared to the explanations given in some of the early 'technology transfer' events of the 1980s, which targeted another element of the industrial audience, the corporate executive and business community. I draw particularly from a book based on a colloquium organised by the MIT Industrial Liaison Program (Winston and Prendergast, 1983), asking both about the terms in which AI was described and how the audience was configured. In 7.2 I look at texts produced by and for the policy making community, asking in what terms AI was represented as a strategic technology. I draw on policy discussions including government publications (such as the Alvey Report, DOI, 1982), industrial conferences (eg SPL, 1983) and some publications addressing a more general audience (eg Feigenbaum and McCorduck, 1983). Finally, in 7.3, I draw the case study together by returning to the question of the failure of AI, as represented by members of the industrial AI community, asking how their diagnosis (AI failed because it was oversold) should be understood in the light of the variety of ways in which AI was 'sold' (as shown in Chapters Four to Seven).

7.1 What computers can do: non-numerical computing for an industrial audience?

I have previously (Chapter Five) suggested that symbolic processing provided a loose but unifying description of AI within the context of cognitive science, and discussed the extent to which this unifying description may have been reinforced by the (temporary) abandonment of neural nets as a possible model of intelligence. Such descriptions and distinctions were invoked within internal discussions, in textbooks for students (new members of the community) and in overviews that enabled the community to reflect on its own position. There is another distinction which the symbolic paradigm served to enforce, and that was the distinction between AI and the rest of computer science. The history of AI can be told as part of the history of computer science, or in opposition to it. For example, Feigenbaum famously[163] told McCorduck the story of his simultaneous introduction to AI and to computing, as a member of Simon's graduate seminar in 1956. Simon announced to the seminar that over Christmas he and Allen Newell had invented a thinking machine. Feigenbaum recalled (McCorduck 1979, p 116):

> And so we said, 'Well, what do you mean by a thinking machine? And in particular, what do you mean by a machine?' In response to that, he [Simon] put down on the table a bunch of IBM 701 manuals and said, 'Here, take this home and read it and you'll find out what I mean by a machine.' ... So we went home and read the manual - I sort of read it straight through, like a good novel. And that was my introduction to computers.

In a later telling of the story, Feigenbaum (1992, p 194) reports that he stayed up all night reading the computer manual, and

> as the dawn came, I rose a born-again computer scientist, though of course the term had not been invented yet, and would not be for another decade.

This is the AI researcher as computer scientist. Most of the public histories of AI, however, emphasise the difference between AI and other approaches to computing. Simon, for example, told McCorduck (p 129) about a visit to an air defence lab in Santa Monica in 1952:

> But that air-defence lab was really an eye opener. They had this marvellous device there for simulating maps on old tabulating machines. Here you were, using this thing not to print out statistics, but to print out a picture, which the map was. Suddenly it was obvious that you didn't have to be limited to computing numbers -

[163] This anecdote is repeated in Simon (1991) and Feigenbaum (1992), in both cases they reference McCorduck (1979).

> you could compute the position you wanted a spot to appear on a piece of paper. You could print pictures with things that weren't even a modern computer, just old card calculators.

AI researchers tended to claim for themselves the insight that computers could do more than crunch numbers as Claude Shannon complained to McCorduck (describing a conference held at MIT during World War II) (pp 100-101):

> You may say [it] isn't artificial intelligence, that it's a different thing, but I see that as everybody trying to find the farthest reaches of computers. We realised that this was a lot more than an adding machine, a much more general and powerful tool than that.

Shannon's complaint testifies to contemporary assumptions among AI researchers that AI constituted a 'break' from numerical processing (and incidentally indicates that there are other histories available).

In *The Handbook of AI*, a distinction between numeric and symbolic computing was introduced at the beginning of the first volume, in explaining the purpose of the book and identifying its intended readers (Barr and Feigenbaum, 1981). The editors first explained the need for such a book in terms of 'bridge building' to other disciplines (including 'our own colleagues in computer science' (p xi)), as a source of texts for college courses, and as a means of educating industrial researchers. They explained (p 12):

> Most scientists and engineers, though very knowledgeable about the 'standard' computer methods (eg numerical and statistical methods, simulation methods), simply had never heard about symbolic computation or Artificial Intelligence.

That is to say, the point of the book was given as explaining symbolic computing to those already proficient in numeric computing. In addition to identifying its readership as being largely from the world of numeric computing, The *Handbook* also configured its readership as one for whom it was appropriate to explicate specific concepts by reference to more familiar computing ideas. An example of this is the explanation of AI programming languages by reference to familiar programming languages. In the introductory overview to the section on AI programming languages, volume two of the *Handbook* (Barr and Feigenbaum, 1982, p 3) explained that, just as Fortran and Cobol provide 'higher level algebraic and business primitives, respectively', so AI programming languages provide higher level concepts for

demonstrating and developing AI ideas.[164] The *Handbook* went on to explicate 'AI ideas' in terms of the manipulation of symbols rather than numbers. It stated (p 3):

> The first and most fundamental idea in AI programming languages was the use of the computer to manipulate arbitrary symbols - symbols that could stand for anything, not just numbers.

At the same time, the article also appealed to the needs of its readers as programmers, and remarked on the excellence of AI programming languages as programming environments:

> Most AI programming languages, in addition to supporting many quite novel high-level features, offer splendid environments for writing, debugging and modifying programs.

The more detailed articles focused on technical explication, in which the more general explanation of what AI is would be out of place, and references to the symbolic paradigm tend to be absent within the technical articles. Nonetheless references to the symbolic are to be found in the introductory explication of a broad range of AI sub-fields,[165] and may be said to provide a unifying description of the field, which also implies the radical novelty of the approach.

Besides describing AI as a radical new approach to computing, the *Handbook* also presented AI as opening up a novel range of applications, primarily expert systems. However, in adducing the use of (or need for) expert systems, the *Handbook* had resource not only to the claim that AI was non-numerical computing, but also to the idea that it was concerned with simulating human behaviour (that is, the distinction between AI and the simulation of behaviour was not insisted on in this context). The authors of the overview claimed (p 79) that the field of expert systems aimed at a general understanding of 'the nature of knowledge':

---

[164] The *Handbook* explains (Barr & Feigenbaum, 1982, p 3): 'AI programming languages have had a central role in the history of Artificial Intelligence, serving two important functions. First, they allow convenient implementation and modification of programs that demonstrate and test AI ideas. Second, they provide vehicles of thought: as with other high-level languages, they allow the user to concentrate on higher level concepts. Frequently new ideas in AI are accompanied by a new language in which it is natural to apply these ideas'

[165] The role of the symbolic in explicating AI programming languages has been mentioned. The symbolic paradigm was used in explications of expert systems (Barr and Feigenbaum, 1982, p 79): 'These systems were designed to manipulate and explore symbolically expressed problems...' and expert systems, are both referred to in this section. Another example is in natural language processing (Barr and Feigenbaum, 1981, p 227): 'The computer, like the human mind, has the ability to manipulate symbols in complex processes, including processes that involve decision making based on stored knowledge. It is an assumption of the field that the human use of language is a cognitive process of this sort.'

> ... both in terms of formal representational systems and as an essentially social phenomenon - knowledge as something that must be shared and transferred between men and machines.

And they implied (p 80) that the goal of expert-systems design was to produce a system which would be quasi-human:

> For an expert system to be truly useful, it should be able to learn what human experts know, so that it can perform as well as they do, understand the points of departure among the views of human experts who disagree, keep its knowledge up to date as human experts do (by reading, asking questions, and learning from experience), and present its reasoning to is human users in much the way that human experts would (justifying, clarifying, explaining and even tutoring).

The need for such systems was given rather broadly in terms of supporting experts, replacing experts and knowledge archiving. In the context of the AI debate (as discussed in Chapter Six), some critics of AI challenged the coherence or feasibility of such claims. However, in the context of technology transfer some slightly different questions arise concerning the need for such a system, or how specifically the need for it is construed in the text. Whereas the reader as computer scientist was argued into accepting AI as a good way of programming, the reader as user of AI was presented with a future that is discontinuous with her present working life, in which computers are co-workers rather than tools. Indeed, even from a computer science point of view, this involved a number of what Feigenbaum and McCorduck (1984, p 156) described as 'scheduled breakthroughs'. One problem with this view of the future is that it was arbitrary, other futures might happen. Another problem is that it failed fully to enrol users in construing new needs. However, the *Handbook* was not addressing users, but systems designers and programmers.

One problem in explaining AI to industry was that industry was not a homogeneous audience. That is, it was not enough to sell AI to industrial systems designers; corporate executives had also to be convinced. This audience became important to the AI community during the early 1980s when technology transfer conferences and books began to be produced, explaining AI to a corporate audience. One early example was a colloquium organised by MIT's Industrial Liaison Program in collaboration with an investment banking firm (F Eberstadt and Company). The papers and edited discussion from this colloquium were later published as *The AI Business. Commercial Uses of Artificial Intelligence* (Winston and Prendergast,

1984).[166] The colloquium included speakers from the MIT AI Lab, from AI vendor companies, from user companies (including Schlumberger and DEC) and from two venture funding companies.[167] Recognition of the specific interests of corporate (entrepreneurial) listeners was signalled in an overview address by Patrick Winston, then Director of MIT's AI Laboratory, which opened with the words:

> The primary goal of Artificial Intelligence is to make machines smarter. The secondary goals of Artificial Intelligence are to understand what intelligence is (the Nobel laureate purpose) and to make machines more useful (the entrepreneurial purpose).

Randall Davis (an associate professor in the MIT AI Lab) warned the audience that expert systems were slow to build,[168] and that very few had yet been fully implemented. He said (p 27):

> On a scale from a gleam in somebody's eye to wide commercial use, five systems have reached the stage of commercial use.

The five systems identified by Davis as in commercial use included two of the earliest, Macsyma[169] and Dendral[170] (both started in 1965); the other three were Steamer,[171] R1 (otherwise known as XCON) and Dipmeter Advisor. XCON (which was developed and used by DEC for configuring computer systems) and Dipmeter Advisor (developed and used by Schlumberger for log analysis in oil exploration) were both the topics of papers given at the colloquium, along with a description of an experimental medical diagnosis system (Caduceus). Arnold Kraft of DEC (p 42) stressed both the regular use and the benefits of XCON:

> We have used XCON daily in our plants since 1980. So far XCON has analysed nearly 20,000 unique orders and is running now with 95 to 98 percent accuracy. It has

---

[166] The AI topics discussed at the colloquium went beyond expert systems, and also included discussion of robotics and some other topics such as natural language front-ends.

[167] The Preface to the book (Winston and Prendergast, p 1984) explains that the purpose of the colloquium was 'to bring together four groups of people: one group to supply the academic perspective, another group to represent the hard core, financially oriented people, a third to represent the industrial research and development people who can look at the questions from both sides, and a fourth to represent solutions-oriented people, who use Artificial Intelligence without admitting it, because there is a job to be done.'

[168] He estimated 'at least five man-years of effort' for 'a substantial expert system' (Davis, 1984, p 26).

[169] Macsyma was designed to assist in the solution of mathematical problems. According to the Handbook (Barr and Feigenbaum, 1982, p 143): 'Macsyma is used extensively by hundreds of researchers from government laboratories, universities and private companies throughout the United States.'

[170] Dendral assisted chemists in identifying molecular structures.

[171] This is described in the Glossary to Winston and Prendergast (1984, p 316) as: 'Experimental instruction system that teaches propulsion engineering.' The qualifier 'experimental' raises a question as to whether it was in real use.

> become an indispensable and effective business tool.

For a corporate audience, the expert system is judged in terms of cost-effectiveness, not 'intelligence'.[172]

A further problem of marketing AI to industry was also illustrated by two presentations at the MIT colloquium, given by representatives of venture funds (William Janeway, from the investment bank Eberstadt and Frederick Adler from Adler and Co), both of whom were sceptical about the value of selling products as AI. Adler (1984, p 258) distanced himself from the very idea that his company had financed AI:

> Someone recently told me that he understood my company put more money into Artificial Intelligence than any other company had. I told him that I did not know what he was talking about. We have not put a dime into Artificial Intelligence.

Adler's point was partly that the companies he invested in were at best 'on the edge of Artificial Intelligence', but partly that investments in AI companies would be based on the same decisions as for any other companies.[173] Nonetheless, his remark is reminiscent of the advice 'Don't tell the client it's AI' (cf 3.3). Janeway (1984, p 266) remarked on a paradox in corporate perceptions of AI:

> Beyond a small number of industrial projects and a comparable number of missionary entrepreneurs, we found widespread disdain of Artificial Intelligence and even more widely spread ignorance among a variety of people who in fact were engaged in projects recognisably similar in purposes and even in approach to artificial intelligence programs. What is AI? We formulated a tentative proposition: Artificial Intelligence ceases to be Artificial Intelligence when it enters the real world, at which point it becomes something like advanced Computer Science.

In this passage, Janeway obliterates the difference between AI and computer science - suggesting that 'AI projects' were at least similar to some non-AI projects - and also appears to reinvent the 'moving boundary' of AI remarked on in 6.2 It is interesting, therefore, to compare Janeway's description of the moving boundary of

---

[172] XCON became something of an iconic case study, as an expert system that was in use and saving DEC money. Eventually most of the computer mainframe manufacturers adopted similar systems. The UK computer manufacturer ICL claimed that its system SD Adviser saved it £5 million a year (*MIN*, Nov 1987, pp 2-3).

[173] In illustrating the commercial questions, Adler (1984, p 260) asks: 'Is the need large enough so it will be reproducible in volume? Can it reach the $50 million figure? Is the team good enough to do it? Is the management profit oriented? If the motive is to bring Artificial Intelligence into the 1980s, or something romantic like that, that is wonderful, but if not done at a profit, it will not happen. It is a waste of time.'

AI with a rather similar description by Minsky (1988, p 307)[174] , made shortly after the remark (quoted in 6.2) that 'intelligence' describes 'the moving horizon of our growing understanding of how minds might work':

> Indeed, from yet another point of view, I sometimes think of AI as 'the current frontier of computer science'. ... Then, in that view, AI is simply finding ways to make computers do the useful things that no one yet knows how to make them do. This lazy comprehensiveness has one annoying side-effect of making AI's cumulative reputation subject to a continual 'exponential decay' - wherein each achievement fades away to be credited to some other speciality.[175]

For both Janeway and Minsky here, the moving horizon is not related to the semantics of 'intelligence'. It is a phenomenon related to the academic discipline and its relation to computer science and to industrial implementation. In Minsky's description, the loss of sub-fields makes AI a field of no achievements (which is a problem in relation to institutional requirements of accountability), but this could perhaps be explained simply in terms of its fast development, a problem of success. Janeway's point (pp 269-271) is slightly different and contains a moral for those selling AI (a *caveat vendor* rather than a *caveat emptor*):

> [I]n the financing of any technology high enough to lie beyond the comprehension of the vast majority of potential investors, the guiding rule should be caveat vendor, let the seller beware, for it will be the seller who will be called to account once the hype has ended ...

It is interesting how the word 'hype' is introduced into the argument. Janeway's advice was to avoid investment in AI *as* AI.[176] However, he deploys the trope of the moving boundary not simply to represent AI as an unproven technology, but as an unprovable technology, a technology that will always be too new to judge and which (therefore) will always involve mis-selling or hype, since once it can be realistically evaluated it will no longer be 'AI'. Insofar as Janeway accurately

[174] The original publication date of Minsky's article was before 1985. I am quoting from a collection of papers reprinted from AI Magazine, entitled Readings from AI Magazine, 1980-1985 (Engelmore, 1988). It is not therefore clear whether Minsky or Janeway has precedence on this particular formulation of the moving horizon. However, the question of precedence does not seem to me to be an interesting issue here; the interest is in the specific differences in how each deploys this formulation.

[175] Minsky (1988, p 307) gives a number of examples of the decay of subfields: 'In AI's early days we were concerned with recognising patterns of many kinds. Today, "pattern recognition" has become a separate field; it has journals of its own, nor will AI journals accept papers on that subject. Similarly a new field of "symbol manipulation" emerged from AI research efforts like our MACSYMA project, now seen as in the field of 'symbolic applied mathematics".'

[176] ' Janeway concluded his paper (p 271) with the words: 'Only some pieces of the future of Artificial Intelligence should be financed [...] and those may be the ones that by definition no longer are Artificial Intelligence.'

reflected the views of industrial investors, these remarks suggest that the AI community failed to retain full control of its audiences in the context of technology transfer: the claim that this was an effective new form of non-numeric computing, capable of cost-effective application, met a sceptical audience.

7.2 The fifth generation as the future of computing

During the early 1980s a number of 'fifth generation' computer funding programmes were set up around the world which identified AI as providing the paradigm for the future development of computing. The identification and promotion of AI as a strategic technology suggests an approach to technology policy which is nowadays often dismissed as 'picking winners' (cf Henkel et al, 1999, p 190). The texts that I look at in this section are all concerned with technology policy but range from books for a general readership (such as Feigenbaum and McCorduck, 1983 (1984)) to policy texts produced by committees for an audience of policy makers and academic researchers. In between are some texts which were accessible to general readers but address rather specialist interests; for example, books produced by individuals with personal knowledge of the Japanese programme (Moto-oka and Kitsuregawa, 1984)[177] or the Alvey programme (Oakley and Owen, 1989)[178] , and conference proceedings (eg SPL Insight 1983; 1984). In this section I ask how AI came to be identified as a strategic technology by policy makers. That is to say, in what terms was AI identified as a strategic technology, who was convinced (who was the configured audience) and what was the context in which the identification of a strategic technology was made? I ask these questions with particular reference to the UK Alvey programme. This is partly because the Alvey programme may be seen as taking place at an interesting moment in the history of UK technology policy, and partly because the discussion round the Alvey programme is relatively well documented and provides snapshots of some of the terms through which fifth generation programmes were argued for.

---

[177] Tohru Moto-oka, professor of Electrical Engineering and director of the Computer Centre at Tokyo University, was chairman of the Fifth Generation Computer Systems Project and the National Project of Scientific Super Computers. Masaru Kitsuregawa was a member of the parallel processing mechanism working group of the Fifth Generation Computer Project.

[178] Brian Oakley was Director of the Alvey Programme from 1983 to 1987 (and previously Secretary of the SRC), Kenneth Owen was a freelance writer and editor of Alvey News, the newsletter published by the Alvey Directorate.

UK science policy in the early 1980s was still governed by the ideas of the 1971 Rothschild report whereby government supported applied research on a contractual basis, and basic research in responsive mode (cf Tisdell, 1981)[179] , and industrial policy was still influenced by the approach introduced by Tony Benn in the 1960s of support for a single chosen company in each sector. In this context, the Alvey Programme is often represented as a significant move towards collaborative research involving multiple industrial and academic partners (Guy and Georghiou, 1991, pp 153-6; Oakley and Owen, pp 97-98). Faulkner and Senker (1995, pp 14-15) note several trends in the period from about 1980, including an emphasis on university-industry links and a growing stress on the exploitation of publicly funded science. Among these trends, they comment on a conflict between, on the one hand, the Thatcherite emphasis on minimising government intervention and, on the other hand, an increasing tendency towards directed (as opposed to responsive) research through strategic programmes such as the Biotechnology Directorate and the Alvey Programme. Oakley and Owen (p 8) trace criticism of responsive mode science funding back to Iann Barron's membership of what was then the Science Research Council (SRC)[180] in 1974.[181] However, they also remark (p 40) that in comparison with funding programmes in the US (such as the DARPA Strategic Computing Initiative which was funded on a contractual basis) and in Japan, the Alvey Programme was not regarded as 'directed'. In 1980, early in the period of the Thatcher government, Kenneth Baker (then a backbencher) called for a national strategy for information technology.[182] Baker presented a ten-point plan concerned with mechanisms for developing a technology strategy and for exploiting the results. His first demand was for a Minister for Information Technology. By 1981, not only

[179] Tisdell (1981, p 130) comments 'Recent developments in British science policy can only be understood by reference to the Rothschild Report'.

[180] The Science Research Council became the Science and Engineering Research Council (SERC) in April 1981. In 1993, following publication of the white paper Realising Our Potential (OST, 1993), the Engineering and Physical Sciences Research Council (EPSRC) was created.

[181] Barron (who went on to found the semiconductor company Inmos) said he was 'appalled' at the procedures of the SRC and told Oakley and Owen (1989, p 8): 'I proposed that instead of what they were doing, which was peer review of arbitrary programmes, there should be some directed research. There should be a specific goal, and the research should be directed by SRC with specific targets. People should be requested to do particular items of research, rather than coming up and saying "We'd like to play in this area".'

[182] Baker presented a paper entitled National Strategy for Information Technology in June 1980 at an Online conference on telecommunications. The paper is said to have been written in consultation with the Chairman of Logica, Philip Hughes (cf Oakley and Owen, 1989, pp 10 ff).

had this position been created, but Baker himself was installed in it.[183]

In the search for a UK IT strategy in 1981, funding for AI might not have seemed a prime candidate, given the legacy of the Lighthill Report (1973)[184] for the SRC. Lighthill criticised research in AI (suggesting, among other things that it was not a unified field) and his report led to funding cutbacks, especially in robotics work led by Donald Michie and Richard Gregory at Edinburgh University.[185] It is worth remarking briefly on the terms in which Lighthill criticised AI, since this involved him describing the field in terms of his own categories, or what he called the 'A, B, C of AI'. On Lighthill's account, A stood for Advanced Automation; C for Computer-based research into the Central Nervous System; and B stood for a Bridge between A and C, and also for Building Robots (I am following Lighthill's own use of upper case here). He suggested (p 13) that A and C were legitimate areas of research (even if likely to have a high failure rate), but that B

> raises doubts about whether the whole concept of AI as an integrated field is a valid one.

Interestingly, much of contemporary work in AI could have been placed under category A, particularly work that was later called 'expert systems'[186] ; Lighthill expressed guarded approval for Heuristic Dendral (p 11) in particular, and more generally (p 4) for:

> combining a well structured knowledge base and an an advanced problem solving capability.

Indeed, when I interviewed him in 1986 (MIN, Feb 1986, p 3), he claimed that his report correctly predicted the fruitful aspects of AI. He said:

> I think my report may have helped by damping down the enthusiasm for a particular way of trying to go forward - by cooperation between biologists and computer scientists.

In the report, he objected equally to claims that AI represented 'another step in the

---

[183] Baker was the second Minister for IT. Adam Butler was the first, lasting only two months (November and December 1980).

[184] The Lighthill Report was presented to the SRC in July 1972 and published by the SRC in 1973, with replies by the AI community.

[185] I remember that in informal discussion, the Lighthill Report was often characterised as a personal attack on Donald Michie. Cf also a remark to that effect by Feigenbaum and McCorduck (1984, p 202).

[186] I am not sure exactly when the term 'expert systems' came into use, but it seems to have been in the late 1970s. It is used in the *Handbook* (Barr and Feigenbaum, 1982); but not by McCorduck (1979), although she discusses Dendral (calling it a 'knowledge-based system') and some other systems that were later known as 'expert systems'.

general process of evolution' (p 14) and to two particular topics within robotics - 'hand-eye co-ordination' and 'visual scene analysis' (p 7) - which are (in 1999) still considered advanced research areas.[187] The tone of his report, however, seems to have been at least as important as its content. Eight years after publication of the Lighthill report, in 1981, a group of UK researchers met to try to 'rehabilitate' AI and to suggest an SERC specially promoted programme (SPP) in AI; John Taylor[188] told Oakley and Owen (p 15):

> We were sitting around a table wondering what to call this new area. [...] Should we call it artificial intelligence? We didn't want to call it artificial intelligence because of all the Lighthill connotations, and we didn't want to call it expert systems, and we came up with this awful phrase 'intelligent knowledge-based systems' or IKBS.

The plan for an SPP was merged into the Alvey programme, and the phrase IKBS survived into Alvey. How was IKBS transformed from a field in need of more support, to the focus of a strategic technology programme? Part of the answer, at least, has to do with the spate of fifth generation programmes that were instituted around the world in 1981. Indeed, it was explicitly claimed in the Alvey Report (DOI, 1982, para 1.2):

> The catalyst to the formation of the [Alvey] Committee was the unveiling last October of Japan's Fifth Generation Computer Programme.

This, however, raises the question of how the Japanese programme acted as a catalyst, or in what terms it was adduced in justification of the Alvey Programme.

The idea that the Japanese Fifth Generation Computer Programme (FGCP) led to other fifth generation programmes around the world is open to a number of interpretations. Feigenbaum and McCorduck in their lobbying book, *The Fifth Generation*, argued that US dominance in computing was under threat from Japan, while conceding that the US was actually some way ahead of Japan. They set the tone in the first chapter (1984, pp 13-14):

> Today we dominate the world's ideas and markets in this most important of all modern technologies. But what about tomorrow?

Moto-Oka and Kitsuregawa (1984, pp 8-10) listed six fifth generation projects around the world which they claimed were a 'response' to the Japanese FGCP: the

---

[187] I am grateful to Paula Gomes for confirming this latter judgement.

[188] Taylor at that time was head of the Command Systems Division of the Admiralty Surface Weapons Establishment and also Chair of the Computing and Communications Subcommittee of the SERC Information Engineering Committee.

Alvey Programme in the UK; DARPA's Strategic Computing Initiative, and an industrially-funded collaboration centre, the Microelectronics and Computer Technology Corporation (MCC) in the US; the European Community's plan for the Esprit project and funding or projected funding in Germany and France. This list of fifth generation projects was reasonably well established, in the sense that it is repeated elsewhere (cf, for example, the proceedings of SPL Insight, 1983, which included presentations from, or descriptions of, all these projects). However, there is some variation in the extent to which those various projects were justified by mention of the Japanese FGCP. The DARPA Strategic Computing Programme announcement (1983), for example, described itself as 'new generation' rather than 'fifth generation' and did not mention the Japanese FGCP. Indeed, DARPA 'insiders' apparently told Feigenbaum and McCorduck (p 303) that the Japanese announcement had 'simply served to sharpen' ideas they already had. The Alvey Report, by contrast, had recurrent references to the Japanese programme, including (as noted above) citing it as the catalyst for the Alvey Committee meeting. It also mentioned the content of the Japanese programme and summarised the aims (paras 1.3; 3.18). However, the Japanese programme was not the only one remarked on by the Alvey report; it cited the existence of programmes elsewhere as a reason for mounting a UK programme (para 2.18)[189] :

> The US, Japan and countries in Europe are now all mounting programmes comparable to the one we propose for the UK. These rival programmes present a serious challenge to the UK, which we must face.

It seems that the number of fifth generation programmes being planned (of which the Alvey Programme was itself one of the earliest) became a justification for proceeding with the Alvey Programme. Indeed, it might be said that the list of fifth generation programmes provided (was available as) justification for each of the constituents of the list.[190]

There was some overlap between the various fifth generation programmes. The common themes included areas that were usually recognised as AI, but the idea of the

---

[189] The paragraph quoted is from the executive Summary. The point was made at more length in Chapter Three of the Report (DOI, 1982).

[190] This may be said also of the Japanese programme, as may be seen from the deployment of the list by by Moto-oka and Kitsuregawa (pp 8-10), as noted above. That the list provided justification for each of the programmes, does not mean that it was necessarily deployed as justification; in particular, it appears that DARPA avoided such a justification for its Strategic Computing Programme (cf the comment by Feigenbaum and McCorduck, p 303, noted above).

fifth generation was also linked (more or less closely) to developments in hardware. Before asking about the terms in which AI was linked to the fifth generation, it is worth looking briefly at how the computing generations were adduced, and how the 'fifth' was formulated. There are some variations in the way that the generations were described and these may be indicated by looking at the descriptions given by Moto-oka and Kitsuregawa (pp 31-32) and by Feigenbaum and McCorduck (pp 31-32)[191] . Both sets of authors agree on the first three generations: the first generation was based on valves or vacuum tubes; the second on transistors; and the third on integrated circuits (chips). Both agreed that, at the beginning of the 1980s, the third generation was just coming to an end and that the fourth generation would be based on contemporary developments in chip technology in which an increasing density of logic elements were integrated on a single chip, known as large scale integration (LSI) and very large scale integration (VLSI).[192] Moto-Oka and Kitsuregawa describe the fourth generation as based on LSI, commenting (p 32) that 'we might say that we are just entering the era of the fourth generation'. Fifth generation computers, they suggested, would be based on VLSI which they dated as coming into full use in the 1990s. Feigenbaum and McCorduck, however, identified VLSI as the fourth generation commenting (p 31) that 'VLSI will dominate during the 1980s'. The fifth generation, for Feigenbaum and McCorduck involved the abandonment of a serial or von Neumann architecture. They said (pp 31-32):

> Instead there will be new parallel architectures (collectively known as non-von Neumann architectures), new memory organisations, and new operations wired in for handling symbols and not just numbers.

Indeed Moto-oka and Kitsuregawa also saw the fifth generation as involving the development of parallel architectures and hard-wiring for symbolic operations. Both sets of authors also described fifth generation computing as involving a software revolution, based on then current work in AI, which they described as knowledge information processing systems or KIPS. The major difference between the two descriptions, then, related to forecasts of when VLSI chips would come into use and is partly an indication of the lead that the US had in this technology. The field of VLSI was a major component of the Alvey Programme, despite a minority on the

---

[191] The coincidence of page numbers is a coincidence, not a typing error.

[192] Adrian Stokes (1986) in his Concise Encyclopaedia of Information Technology, describes LSI as 'A technology for constructing computer components in which a high packing density of logic elements is achieved on a single chip (of the order of 10,000 per square inch).'

Committee who thought that, as a enabling technology for IT in general (and not just for AI), VLSI had no place on a fifth generation programme (cf Oakley and Owen, p 284).

All the fifth generation programmes included constituents that were described as AI. The Alvey programme was based on four 'enabling' technologies of IKBS, VLSI, software engineering and MMI[193] or man/machine interface. The Alvey Report (DOI, 1982, para 3.6) said:

> We see these priority areas as basic enabling technologies. We have had little difficulty in identifying them or the associated infrastructure and systems which link them. Necessary for any electronic based activity is secure access to world class software tools and technology together with the design tools and technology for Very Large Scale Integration (VLSI). Also essential for IT is a leading edge knowledge of handling information - especially what is now developing as Intelligent Knowledge Base Systems (IKBS) - and of the interaction of man with machine (MMI).

While the choice of areas is represented as obvious and consensual, Oakley and Owen (1989, pp 38-42) report some of the criticisms and disagreements expressed among the policy making community about this choice of four areas. These included criticism (from Philip Hughes of Logica and Ian Barron of Inmos) about the failure to include a programme devoted to computer architectures so that, as Hughes put it, 'the work on architecture subsequently had to be cobbled together by stealing from other programmes'. That is, discussion among members of the policy making community[194] (in committees or elsewhere) involved difference and disagreement, but the statements in the report suggest consensus and unity. While this may again suggest that the difference is audience, with a unified position deployed in a public document (cf Chapter Four), it is also worth considering the institutional purpose of government reports, that is, to recommend policy. A report that *reflected* differences of opinion (as opposed to claiming to have satisfied diverse opinions) would fail as a decisive policy recommendation, and in addition would lack authority, would fail to inspire confidence and would give ammunition to opponents and

---

[193] Now more usually described as Human Computer Interface or HCI (to avoid the gender-specific connotations of 'man') - however MMI was the name used within the Alvey programme, and I therefore retain it.

[194] By members of the policy making community I refer primarily to those who were involved on committees but more broadly also to those whose views were sought through various consultation mechanisms. The Alvey Committee included individuals representing academia, industry and government.

Treasury officials for arguing against its implementation. In this context, terms which come to represent a consensus are likely to be further strengthened by repetition. Just as the list of fifth generation programmes could serve to justify the individual programmes, so the selection of the four programme areas by the Alvey Committee strengthened the authority of each one of them. This may be seen in the report of a workshop sponsored by SERC and the Department of Industry (SERC-DOI, 1983)[195] to begin developing a programme in IKBS, following the recommendations made by the SERC proposal for an SPP, the Alvey Report and another IKBS review meeting organised by the SERC in 1982. After identifying future developments in IT as being driven by innovations in software rather than in hardware, the report (p 3) went on to state

> The foundations for ways round these problems,[196] and hence for the next major advances in computing, have been laid by research in artificial intelligence, software technology and man/machine interaction, coupled with work on novel approaches to computer architectures. [...] The term 'intelligent knowledge based systems' (IKBS) has been coined to denote the first main generation of these 'semi-intelligent' systems that will emerge during the next few years.

In the next paragraph (pp 3-4), the report further underlined the significance of IKBS:

> [...] The likely significance of IKBS for all industrial nations has been pointed out recently by a number of major initiatives including the Japanese Fifth Generation Computer Project, the Alvey Report, the SERC IKBS proposals and the EEC Esprit programme. All these agree that IKBS will become an increasing major part of Information Technology as a whole and IKBS are therefore likely to be of enormous economic significance for all industrial nations.

It seems that a consensual authority had become attached to the idea that AI (IKBS) represented the future of computing.

---

[195] The workshop, known as the IKBS Architecture Study, was chaired by John Taylor and sat between December 1982 and May May 1983.

[196] The report (p 3) had identified a 'software crisis' in 'limitations inherent in the current "conventional" approach to specifying, designing and programming complex systems'.

7.3 The importance of not being extreme

If the policy discussion was, as I suggested in the previous section, driven by consensus and an accruing consensual authority, there remains a question whether influences from the surrounding, non-policy discussions of AI were drawn on in policy discussions to justify and explain AI. In particular, since it is one of the alleged causes of the supposed problem of overselling AI, did the question of machine intelligence play any part in these discussions? On the whole, policy discussions tended to eschew these questions in preference to technical descriptions and appeals to the authority of academic opinion.[197] However, some broad questions about the value of fifth generation computing were raised in the context of a House of Lords Select Committee on the European Communities into the ESPRIT programme (HOL 1985). On the first day of evidence (28 June 1984), the first witness was Brian Oakley, Director of the Alvey Programme. The Earl of Halsbury[198] (p 3) asked for explanation of the aims of the fifth generation computing, complaining:

> In the early days of computers a very few very distinguished people could specify on a sheet of paper what they were trying to do [....] and always there was a concrete something at the end of the road they were trying to build, a piece of control software they were trying to write, and so on. I find the difficulty is in assessing what is going on in the extreme abstraction with which all these schemes are ascribed. [....]

Oakley replied:

> Underlying perhaps all of what are called the Fifth Generation projects is a certain relatively limited number of objectives.

The first that he mentioned was 'conquering the problems of parallel computers', which he discussed briefly. He then continued:

> But the excitement lies in the new areas which are now beginning to become practical, of which the most important probably is what is known as inference

---

[197] Of the texts discussed in the previous section, the only exception to this is Feigenbaum and McCorduck's (1983) book, but this in any case is somewhat orthogonal to the policy discussion (and one way of approaching the question might include asking whether and in what terms Feigenbaum and McCorduck's book is adduced within policy discussions).

[198] Lord Halsbury had previous experience of the field of AI. He had in 1963 been in charge of setting up a specialised computer board in the predecessor organisation to the SRC, the Department of Scientific and Industrial Research (DSIR). Fleck (1982, pp 183-184), who had interviewed Halsbury, claims that this board gave early support to AI following a report by Donald Michie (just returned from a visit to the US): 'Michie's report indicated a widespread positive assessment of the potential of AI among young computer scientists and this undoubtedly formed the basis for the proportionately generous funding by the computing broad of the SRC for research in the area during the mid to late 1960s.'

> computing which differs from normal computing, which is based on the laws of arithmetic and [B]oolean algebra by being more like the way the human being reaches decisions, where the human being thinks of the problem, probably has some degree of abstraction of rules for dealing with that problem, even if the human does not realise that he weighs up a whole number of uncertainties because the chances are that the 'facts' on which he is trying to make a decision are far from being facts, and finally, infers what the right thing to do is. Well, it is now becoming possible through the developments of the Artificial Intelligence community to tackle such problems in computers. There is no question that this will influence the whole development of society. Computers will find a vastly increased range of applications through this type of computing.

This paragraph seems to be making a claim about the possibility of machine intelligence, based on a theory of human reason, and using it to predict the future of computing. Under further questioning, Oakley also suggested that human-like computers would avoid the mistakes of humans,[199] but conceded that professional knowledge might be more difficult to encode than the knowledge of technicians. A hostile critic might identify a danger of exaggeration. A member of the committee, Lord Kearton, remarked:

> I agree very much with the tenor of Mr Oakley's answer. It seems to me to be at an entirely different level than this rather vague talk of artificial intelligences.

Oakley responded to Kearton:

> I have to say, that I believe that the artificial intelligence community, though it has done us a great service in pushing forward the techniques, has done us a great dis-service in trying to draw close parallels with the way human beings work. There is no doubt that we are vastly inferior in our machines to the capabilities of even the most uneducated of human beings.

What is interesting in this interchange is that, although Oakley initially did nothing to pre-empt hostile criticism (cf the practices described in 6.2), Kearton effectively invited him to do so. Oakley then distanced himself from what he implies is a more extreme position of drawing 'close parallels with the way human beings work', and assigned that position to 'the artificial intelligence community'. One of the points of interests of this sort of interchange is the suggestion that what is to count as an exaggerated claim need not relate to the content of the claim, but to the way in which

---

[199] He said (p 4): 'One of the joys is that if the laws of arithmetic are right, then, in general, computers will reach the right answers, providing one looks after small problems like the rounding errors.'

it is presented: as if disclaiming exaggeration is a means of avoiding it.[200] In this respect, the above interchange is similar to the remarks by Dr Payne quoted in 6.2 (p 119) where, as I suggested, a preliminary disclaimer is used to pre-empt accusations of extremism.

One of the implications of the position adopted by Oakley (that AI machines are intelligent like humans but not nearly to the same level) is that it enables intelligent machines (machines that are as intelligent as humans) to be represented as the ultimate or extreme case. This may be compared to explanations of expert systems in the IKBS Architecture Study(SERC-DOI, 1983, p 3), which represents the current generation of IKBS or expert systems as 'first generation' and 'primitive':

> The term 'intelligent knowledge based systems' (IKBS) has been coined to denote the first main generation of these 'semi-intelligent' systems that will emerge during the next few years.
>
> Primitive examples of such 'applied AI' systems already exist as commercial systems doing not-trivial tasks, in particular the so-called 'expert' systems.

Here the importance of IKBS is represented, not simply in terms of what they may be shown to be doing now, but as a beginning: a 'first' generation implying N succeeding generations; a 'primitive' system implying more sophisticated ones. In this case, without invoking claims of 'intelligence' (except in the name, IKBS), a trajectory is implied. In Chapter One, I briefly discussed the argument (Nelson and Winter 1982; Dosi, 1982, 1984) that 'technological trajectories' may be understood as beliefs held by the developers of technology. Nelson and Winter (p 259) also point to the pervasiveness of the idea of technological trajectories in the economic literature:

> [I]n any era there appear to be certain natural trajectories that are common to a wide range of technologies. Two of these have been relatively well identified in the literature: progressive exploitation of latent economies of scale and increasing mechanisation of operations that have been done by hand.

The idea of a trajectory, whatever its faults as an explanatory mechanism (cf MacKenzie, 1992, pp 30-35), is one that is familiar in discussions of technology in the economic and policy literature. There is, as MacKenzie suggests, an implication of technological determinism in the concept of technological trajectory, an

---

[200] This is quite a common form in ordinary speech - but can act as an excuse as much as a disclaimer, eg, 'I don't want to be rude but ...'

assumption that the trajectory *will* be followed. My interest here, however, is how the practice of representing a technology in terms of a trajectory, acts to situate that technology as having a progressively important future - as the Architecture Study does in the case of IKBS. Further, as the remarks by Oakley demonstrate, the assumption of a trajectory may in effect also be applied to the simulation of intelligence in such a way as to both enhance the importance of IKBS (which may be represented as more intelligent than conventional systems) but at the same time attempt to disarm a hostile critic (ie that this claim is not 'extreme', it is not claiming that the system will be *as* intelligent as humans).

The trajectory of intelligence - seen as *the* future of computing - also relates to the construction of user needs. I noted above, in 7.1, that when it came to construing user needs, *The Handbook of AI* (Barr and Feigenbaum, 1982, p 80) represented a useful system as a human-like system. DARPA's Strategic Computing Programme was also based on the assumption that human-like computers would be useful. In the programme announcement (DARPA, 1983, p 1) it claimed:

> In contrast with previous computers, the new generation will exhibit human-like, 'intelligent' capabilities for planning and reasoning. The computers will also have capabilities that enable direct, natural interactions with their users and their environments as, for example, through vision and speech.
>
> Using this new technology, machines will perform complex tasks with little human intervention, or even with complete autonomy. Our citizens will have machines that are 'capable associates', which can greatly augment each person's ability to perform tasks that require specialised expertise. Our leaders will employ intelligent computers as active assistants in the management of complex enterprises. As a result the attention of human beings will increasingly be available to define objectives and to render judgements on the compelling aspects of the moment.

Similarly, in describing the purposes of the Japanese FGCP, Moto-oka and Kitsuregawa (1984, pp 80-82) claim:

> The first thing that users can expect from fifth generation computers is that they will be easy to operate. To achieve this, not only will the human-machine interface have to be 'near human' in nature, but the computer itself will also have to possess 'common sense', ie common knowledge on a par with that possessed by humans.

Further:

> The second capability that users can expect from fifth generation computers is that

> they will gradually assume the monotonous, boring jobs now being done by humans, thus freeing the user to devote him/herself to more complicated and challenging tasks.

The construed needs (human-like computers) and the assumed trajectory (increasingly intelligent computers) therefore seemed to support one another. Interestingly, insider critics of AI, and others in the business of advanced systems design, frequently attacked the construed *needs* of the AI story, arguing that what was required was better tools not more human-like systems (cf Cooley, 1980; Göranzon and Josefson, 1988; Rosenbrock, 1989; Winograd and Flores, 1986).

The question nonetheless remains whether the policy makers of the fifth generation computing programmes (or the AI research community before them) can justifiably be blamed for overselling or misdescribing AI as the future of computing in the 1990s. From the perspective of the late 1990s, this is not a straightforward question to answer. If the question concerns the development of 'intelligent' computers, then we may look around and not see any; however, this was only to be expected from what we have learned about the moving horizon of ascriptions of intelligence (cf section 6.2; Minsky, 1959; 1988; Woolgar, 1985; 1989). If the question concerns the application of AI, then again we may look around and not immediately see AI in use. However, it may also be argued that there are a number of socio-rhetorical mechanisms at work acting to render AI invisible. These include another case of a moving horizon, this time the moving horizon of AI (cf section 7.1; Janeway, 1984; Minsky, 1988), as well as the prevalence of deliberate marketing strategies not to tell the client it's AI (cf section 3.3; 7.1). If the question concerns the application of expert systems (IKBS), then the answer may refer either to the perceived limitations on expert systems (cf the remarks by Patric Taillibert reported in 3.3) or to widespread use of rule-based or knowledge-based systems as one more tool for systems designers. It's worth also noting in respect to both these answers, that expert systems are no longer 'AI' - that is, they are no longer deployed in tasks considered AI (which are mostly connectionist projects), and where they are used, they are not considered AI. This may be illustrated by the following paragraphs from the FAQ (frequently asked questions) section of IBM's Deep Blue web page, put up following Deep Blue's victory over chess champion Gary Kasparov in 1997:

Does Deep Blue use artificial intelligence?

No .....

Its strengths are the strengths of a machine. It has more chess information to work with than any other computer, and all but a few chess masters. It never forgets or gets distracted. And it's orders of magnitude better at processing the information at hand than anything yet devised for the purpose.

'There is no psychology at work' in Deep Blue, says IBM research scientist Murray Campbell. Nor does Deep Blue 'learn' its opponent as it plays. Instead, it operates much like a turbocharged 'expert system', drawing on vast resources of stored information

What is interesting about this interchange is that the answer 'No' is justified by a description that includes the system's access to 'information' and its similarity to an 'expert system'. More generally, it may be said that the technology of AI (as described for example in *The Handbook of AI*), that is, the programming languages (such as Lisp and Prolog) and paradigms (such as rule-based programming and object-oriented programming) continue to be used, although they are not (as the fifth generation programmes claimed they would be) at the strategic heart of computing in the 1990s. However, it is also often argued that the contribution of AI to computing cannot be measured simply by what counts as AI, but that it has had a hand in most aspects of computing advance (cf for example, Minsky 1988). Many specific predictions of the fifth generation programmes are difficult to evaluate, such as the claim that by the 1990s personal computers with human-like interfaces would be in widespread use. Is that true or false? The PCs on our desks in the 1990s are not based in Lisp or Prolog as the fifth generation programmes predicted. However, we do mostly have personal computers on our desks, which was a radical prediction in the early 1980s. Do they have human-like interfaces? Well, at least one document explaining the Japanese FGCP (Stewart, 1985, p 55) identified the Apple Macintosh as the first AI computer:

The Macintosh [...] is designed from the bottom up to handle object-oriented programming, making it a perfect environment for 'windowed' software, 'icon' languages and graphics in general.

Windows, icons and graphics are all familiar on our desktops, all based on work from Xerox Palo Alto Research Centre (PARC), which Stewart (p 55) describes as 'one of the seminal centres of AI research', first taken up by Apple then reproduced in Microsoft Windows. This might be said to be a good reason for describing familiar

computer systems of the 1990s as 'AI' and 'fifth generation'. At the very least, these interfaces have acted to make DOS line-command interfaces look like something from a previous generation.

Where then, was the hype? In the variety of predictions associated with the AI research community and fifth generation computing programmes, it is possible to identify some that seemed particularly outrageous when made; and some that may be argued to have come about and some that may be argued to have failed. Perhaps 'hype' is to be identified in the provocative style adopted by some AI researchers (cf 6.3); however, this tone was avoided in technology transfer contexts, where AI was justified through academic authority and national expediency (cf 7.2); and it was also avoided in the industrial marketing of AI (cf 3.3; 7.1). Perhaps 'hype' is to be identified in the outrageousness of particular claims, but then these claims have had variable outcomes: the claim that a computer would one day beat the world champion at chess, for example, was for many years considered one of the most outrageous claims (and chess machines were one of the targets of Dreyfus (1979) attack on AI). Now that it's happened, this claim cannot be considered 'hype'. For the critics, in the context of the AI debate (cf 6.1), the claim that computers would be (or already were) intelligent was an example of 'hype'; however, this accusation can be partly dissolved in substantive differences and in differences in disciplinary interest (cf 6.3); more importantly, claims of machine intelligence were largely suppressed in the context of technology transfer (cf 3.3; 7.1). Perhaps 'hype' is located in the melodrama of some of the public histories (cf 4.1); but again this tone was avoided in marketing AI.

If the hype seems always to disappear (at least as an agent of 'the problem of AI'), it may be more fruitful to notice the way in which speakers positioned themselves in relation to 'hype', that is to say, at a distance from it. It might be said that 'the hype is always elsewhere', echoing an observation by Woolgar (1985, p 563) that, according to the responses given by scientists to ethnographic visitors, 'the science is always elsewhere'. However, while it may be odd to find scientists locating the science as elsewhere, it is not so odd to find speakers distancing themselves from hype. The interest is in the way in which locating the hype as elsewhere acts to

signal the measured judgement of the speaker. This position was all the more available in the case of AI because of a history of provocative statements by some members of the academic AI community - starting, perhaps, with McCarthy's choice of the name 'Artificial Intelligence' which failed to be 'sober and scientific' (cf 4.1), as well as the exotic scaremongering of some the narratives for general readers, from early stories of 'electronic brains' to AI claims that computers would be our evolutionary successors (cf 6.2).

Earlier in this section, I quoted a remark by Oakley (HOL, 1985, p 4), in which he made such a distancing move:

> I have to say that I believe that the artificial intelligence community [...] has done us a great dis-service in trying to draw close parallels with the way human beings work.

I previously suggested that this move (like the one made by Dr Payne of Decca Radar (NPL, 1959, p 31; cf 6.2)) involved the pre-empting of criticism through the disclaiming of a reading that might otherwise be made - that is that Oakley was trying to draw close parallels with the way human beings work. Now, however, I am making a further point to do with the performative proprieties of technology transfer discourse, where distancing oneself from hype acts to justify the speaker's authority, contributing to a tone of sober and realistic assessment. A similar move was made, for example, by the investment banker William Janeway (1984) when (as I noticed in 7.1) he identified AI with hype and recommended avoiding investment in AI *as* AI.

To summarise my argument in this section: in the history of explanatory practices associated with AI, invocations of 'intelligence' and of similarities between humans and computers have played a recurrent role; however, in the context of technology transfer this history has been deliberately marginalised, not only through avoidance of the terms, but through identifying them as 'hype' and locating them as elsewhere. This suggests that it cannot straightforwardly or crudely be argued that exaggerated claims of the computability of intelligence led to exaggerated predictions of the power of AI and fifth generation computing which led to the undermining of the AI industry when these claims and predictions could not be met; further I also suggested that most cases of claimed 'hype' could in any case be argued. This does not entirely rule out a more sophisticated version of the argument; I have for example noted that 'intelligence' was sometimes treated as a trajectory, so that the simulation of

intelligence could simultaneously be represented as 'hype' (and as elsewhere) *and* as the direction of development. However, a more sophisticated version of the argument would involve recognising that communication between discursive communities is based in the performance of social relations, and is not a case merely of the passing of knowledge between social groups. This is the thesis that I have been arguing throughout the case study, and I provide a detailed summary account of the arguments in Chapter Eight.

7.4 Summary

In this chapter I have looked at explanations of AI in the context of technology transfer, that is, explanations which primarily address an audience of industrial customers and political funders. In section 7.1 I suggested that the industrial audience was not configured as homogeneous: that industrial researchers were addressed as fellow-researchers and given a privileged (technical) explanation; while corporate executives who were being asked to invest in the technology (primarily by adopting it) were given explanations in terms of needs (applications) and in terms of 'proven' examples of implemented systems. The corporate audience, on this account, appeared to be (and was configured as) sceptical and conservative. By contrast, explanations of AI for a policy-making audience (7.2) adduced that same conservatism as a problem for a nation that needed to stay ahead of, or as close as possible to, the competition. In this context, explanations of AI tended to rely on citing academic authority and asserting (and thereby revalidating) a consensus on the strategic importance of AI as the 'fifth generation' of computing. Finally (7.3) I suggested that for this audience the idea of machine intelligence and the possibility of simulating human intelligence was not frequently invoked; and usually invoked to be denied; and that this undermines the crude argument that academic hype was responsible for undermining the AI industry. The more general purpose of this chapter, as in the earlier chapters of this case study, has been to challenge the idea that technology transfer involves the passing of knowledge which in principle, if not in practice, is available in some neutral, single narrative for all audiences. I present an over-arching review of the case study in the final chapter (Chapter Eight).

CHAPTER EIGHT.
ANALYTIC AND THEORETICAL CONCLUSIONS

8.0 Introduction

In this chapter I conclude my thesis with an overview of the arguments and material that I have presented. I begin (in 8.1) by restating the issues raised in chapters One and Two in the light of the examples and interpretations offered in the course of the case study. In 8.2 I recap the issues and conclusions presented in the case study and then draw out a number of relevant themes. Finally, in 8.3, I ask to what extent and in what way the findings are generalisable, and also what outstanding questions are suggested for further research.

8.1 The analytic and methodological framework

In Chapter One I began by noticing the variety of different disciplinary interests apparent in discussions of technology transfer and drew out two main themes: the first, associated particularly with the school of evolutionary economics, concerned ways of theorising the impact of technological innovation and stressed the uncertainty of technological outcomes; the second, associated with science and technology policy studies, included an interest in modelling links between social groups and in the movement of knowledge, and suggested an unresolved question concerning the local appropriation of knowledge. I noticed that in the case of both these themes it is possible to draw parallels with discussions in science and technology studies (STS) in relation to an interest in the uncertainty of technological outcomes. In the light of discussion in Chapter Six (especially section 6.1), it is relevant to emphasise some differences in disciplinary interest apparent in the way that the three disciplinary fields approach the question of uncertainty in these discussions. For evolutionary economics (cf section 1.1) the uncertainty of technological outcomes is both a problem for economic actors (the impossibility of making a fully rational choice) and an important level of explanation in the macro-economic model, where the behaviour of firms in attempting to manage uncertainty is adduced as a cause of discontinuities (economic growth or collapse). In STS, by contrast, the uncertainty of technological outcomes has been explained (Pinch and Bijker, 1989) in terms of the power of users to determine the meaning of artifacts (what they are and what they are for) and provided an important

contribution to an ongoing sociological debate concerning the relationship between the social and the technical. In attempts to model technology transfer as the flow of knowledge between social groups (cf section 1.2), differences in local uptake of knowledge may be understood in terms of interpretive flexibility and provide a possible contribution to understanding and managing problems associated with different local appropriations of knowledge. From an STS perspective, the questions generated by this overview of issues may be summarised as relating to how actors judge technological and scientific products (in a broad sense of product)[201] and how these products are appropriated locally - an issue that in this thesis I relate to questions of *reading*. In approaching these questions in terms of discussion within STS, it is possible to further criticise the 'social shaping' approach (mentioned above) as making an unwarranted assumption about the priority of the identity of users as social groups. In particular, arguments by Latour (1987) and Woolgar (1991) point to the creation of the user in the design of technological artifacts. However, this risks understating the persistence of local interests. I concluded Chapter One (in section 1.3) by suggesting that what is needed is a way of theorising social groups which allows for the co-production of artifact and user, as well as the persistence of local interests. I proposed that conceptualising technology transfer as involving communication between discursive communities might provide a way of remaining alert to the co-production of community, artifact and audience (user).

My proposal to study technology transfer in terms of communication between discursive communities already implied some methodological parameters - in particular that I was committed to a discursive analysis - and I explored the main issues and some of the specific detail of this methodological approach in Chapter Two. I showed (in 2.1) that ideas of performativity of texts are present through a number of approaches to STS (and in related disciplines), and argued that this provided a way of studying how discursive communities are performed through the negotiation of boundaries and through the co-production of community, audience and technology. In section 2.4 I explored in more detail what this methodological approach implied for my selected case study (which I had introduced in section 2.2) and in the light of my ethnographic situation as an 'indigenous commentator' (discussed in section 2.3). I stressed that in my case study I was looking at the exploitation of Artificial Intelligence (AI) primarily as an example of technology transfer, rather than as a substantive claim about the possibility

[201] By a broad sense of 'product' I mean to include not only exploitable technological artifacts, but also new scientific knowledge, skills and the users and needs configured in the artifacts, knowledge and skills.

of machine intelligence. It is evident from the case study (especially Chapter Six) that the topic of ascribing intelligence to machines could not be entirely ignored, even in a discussion of technology transfer. However, my approach to the topic still maintained a distance, insofar as I discussed the ascriptions as explanatory practices, asking how they acted both to perform community and to configure an audience. This study produced some interesting material about both the explicit and implicit management of audiences which I discuss further below. The discussion of my own ethnographic situation (relevant because the case study is based on material gathered when I was working as a specialist journalist) showed that the methodological issue of ethnographic distance was itself related to questions of the discursive performance of community. This became particularly evident in Chapter Three, where I showed how the broad industrial AI community included a range of indigenous commentators who contributed to performative explanations of AI. In section 2.4 I introduced the terms in which I would approach the task of studying communication between discursive communities as the study of explanatory practices, interpreted in relation to location, context and audience. I gave initial descriptions indicating that I took 'location' to refer to the social situation of the author of the text (ie as vendor, academic, etc); I took 'context' paradigmatically to refer to the context of publication or production, but indicated that there might be further relevant senses, such as historical context; finally I took 'audience' to mean both intended and configured audience, where the intended audience referred to the author's or speaker's intention and the configured audience to the rhetorical mechanisms of capture. This distinction was exemplified in my discussion (in section 6.2) of the difference between the unintended hostile reader, who was nonetheless configured in attempts to pre-empt her criticism.

8.2 Major conclusions of the case study

In this section I recap the issues explored and the conclusions reached in the case study, chapter by chapter, as a preliminary to identifying some critical themes and issues that require to be made more explicit.

The case study began (in Chapter Three) by looking at the explanatory practices of the industrial AI community, that is, how AI and in particular the AI market were explained and described in the period following its launch on the market (the 1980s and early 1990s). This chapter focused on explanations of the market, showing how

problems of the market were adduced in different contexts, but usually in terms that suggested a strategic solution. At the same time, it addressed the question of performance of community through the relationship of indigenous commentators to their audience. In 3.1 I looked at a series of market reports, showing how early predictions of rapid growth gave way to recognition of a crisis, but a crisis that was represented as a bringer of realism and maturity in the industry (allowing further predictions of growth). Similarly, in 3.2 I traced the strategic manoeuvres of the AI vendor companies, as represented by the AI newsletters, as they first abandoned the AI programming language Lisp in favour of the 'conventional' but widely installed language C, and eventually abandoned or marginalised the AI marketplace itself. These strategic manoeuvres were accompanied by diagnoses of what had gone wrong (discussed in 3.3), and various assignments of blame, to the vendors (for overselling), to the customer (for being frightened of innovation) and more rarely to the technology (for being inadequate). Blaming the technology was not, however, common among vendors since it was difficult to distance their products from the technology. By the early 1990s, vendors frequently deployed the accusation of overselling, which could be used to make a distance between the immaturity of the early days of the market and claims of a mature and realistic industry that could now identify and avoid the early faults. I concluded Chapter Three by suggesting that the perceived problem of the AI market and its diagnosis in overselling contained an implicit theory of technology transfer as the passing of accurate descriptions between social groups. The diagnosis and the implicit theory of technology transfer therefore served as a starting point from which to question the history of explanatory practices associated with AI, asking whether, where and to whom AI had been oversold, and whether this could be linked to the alleged failure of AI. The theory implicit in the question was also to be set against my own working assumption that technology transfer involves the coproduction of community, audience and explanation.

In Chapter Four I began (in 4.1) by looking at the early history of AI, in a study which contrasted the history of the field as told to a general audience with the more factional and divisive histories told to peer audiences. I noted that the public history told a story of destiny in which the birth of AI (in 1956) was marked by the coming together of the name, the main actors and the first AI programme. While this history (as exemplified by McCorduck's influential 1979 book) was robust in public contexts and was approvingly cited by members of the academic AI community, it was effectively denied by the stories told to peer audiences by the same researchers who acted as McCorduck's

witnesses. In 4.2 I described some of the variations available in histories told by AI academics for a peer audience, in which there were differences of opinion on the starting point of AI and the identity or significance of 'the first AI programme'. In addition, I noted that in explanations of AI for a peer audience, a distinction was usually introduced between AI and the simulation of behaviour which cut across the unified field described in the general histories. In 4.3 I argued that it should not be concluded that the public histories distorted because they simplified, nor that the peer histories presented the differentiation of accurate detail. The variation in the histories was not entirely explicable in terms of being easy to understand, and histories for a peer audience did, in some contexts, provide a unifying narrative.

In Chapter Five I explored the performativity of historical narratives, in the context of relations between AI and other disciplines which might be represented either as allies or as competitors, in respect both to intellectual authority and access to research funding. In section 5.1 I showed how an interdisciplinary alliance, under the title Cognitive Science, provided a context in which AI was represented as united around the symbolic paradigm (or the Physical Symbol System Hypothesis, or PSSH); I argued that while (by the 1970s) this provided a way of representing AI as a unified field, in contrast to earlier habits of distinguishing between AI and the simulation of behaviour (cf 4.2), nonetheless, distinctions remained available between AI and Psychology within Cognitive Science, and between a literal and a looser understanding of the PSSH within AI. In section 5.2, I discussed a 'revisionist' history which denounced the 'official history of AI' as acting to promote a faction (a faction named in the revisionist history as 'symbolic AI'). I argued that the revisionist history, the history of connectionism, might itself equally be exposed as factional, and that it was indeed a tool of whig history to disguise factional interest as neutral truth. I concluded, however, by pointing to the importance of a whig history in sustaining the identity of a discursive community.

Chapter Six may be seen as addressing issues of reading (including attempts by the author to control readings), which complements the emphasis on narration in Chapters Four and Five. In this chapter I addressed the question which is often taken to be at the heart of the identity of AI: the claim that 'intelligence' may be ascribed to machines in the same way as to humans. However, I avoided the substantive question and looked instead at some of the socio-rhetorical mechanisms revealed by the claims and counter-claims concerning the computability of intelligence. In 6.1, I argued that the AI debate,

as an interdisciplinary dispute (between AI researchers and philosophers), provides examples of the way that disciplines perform community through the reading of texts, displaying and maintaining disciplinary standards in the identification of issues and answers, as well as practices of rigour in exploring those issues. My main example was the way in which the two disciplines differed in their readings of the Turing test, which for AI researchers was primarily read as a functional or operational definition, but for philosophers was taken to be significant as an explanatory example of philosophical functionalism. I suggested that this pointed to the difficulties inherent in communication between disciplines as discursive communities. In section 6.2, I looked at ways in which AI researchers invoked claims about intelligence in explaining AI to different audiences. This included examples of attempts to manage a perceived difficulty, first in defining intelligence, and then of the 'slipperiness' of intelligence as an attribute (identified in STS as a case of interpretive flexibility). I argued that invocations of intelligence in explaining AI to different audiences may be seen as attempts (not necessarily conscious ones) to manage the reader through the exploitation of interpretive flexibility, together with differences in narrative register; however this may be sabotaged by the ultimate freedom of the reader, who may always read the 'wrong' text or read it in the 'wrong' way. I concluded by suggesting that accusations that the AI community wilfully exaggerated machine intelligence cannot be straightforwardly accepted, but may themselves involve a contest of inter-disciplinary authority.

In Chapter Seven, I returned to the specific question of the market exploitation of AI (which I suggested in 2.4 might be identified as technology transfer, narrowly understood) and practices of communicating research products to an industrial and political (funding) audience. In 7.1 I argued that the industrial audience or customer for AI was not homogeneously configured, but addressed either as a technical colleague or a corporate executive. For industrial researchers, an explanation in terms of programming technique was appropriate. However I argued that, for this audience, representations of the use or need for applications of AI fell back on talk of human-like systems. The corporate executive or financier, on the other hand, required to be convinced of the usability and cost-effectiveness of AI, preferably in terms of proven or already implemented systems. This audience was adduced as sceptical and conservative, comparable to the industrial audience configured by AI vendors and systems designers as not to be told 'it's AI' (cf 3.3). In 7.2 I looked at ways in which AI was explained and justified as fifth generation computing, by and to policy makers, in the context of

national and international information technology strategies in the early 1980s. I argued that in these texts fifth generation programmes were justified as strategic (and the future of computing) partly by an appeal to academic authority and partly by citing the existence of other fifth generation programmes, so that the list of fifth generation programmes acted as a self-justifying consensus. Finally, in 7.3 I asked whether claims of machine intelligence had been adduced in explanations of AI in the context of policy discussions and for policy making audiences. I suggested that while such explanations were available and sometimes deployed, they were likely to be accompanied by disclaimers that down-played the claim in making it - a move also sometimes made by industrial researchers. I suggested that these disclaimers might also be deployed to imply a trajectory, where 'extreme' claims of intelligence might both be dismissed as hype and represented as the ultimate goal. These disclaimers of hype concealed the need to construct user needs in the marketing of AI on the assumption that human-like systems were the only future of computing. Finally, I concluded that the diagnosis of hype or overselling as the problem of AI, with its implied ideal of technology transfer as the accurate passing of neutral knowledge (cf 3.3), was too crude; rather, complaints of the overselling of AI were found at least from the earliest moments of attempts to exploit AI (in the early 1980s), and that these complaints acted then (as later) to present the speaker as a critical and credible judge of the technology. More generally, the case study provided evidence of the way in which social relations pervade communication between discursive communities; and this is one of the main issues to be emphasised in the following section.

8.3 Contributions to the discussion of technology transfer

In this section I draw out the significance of this thesis for the discussion of technology transfer. I first consider the status of my case study, as an interpretive analysis, before addressing the contributions made by my claim that communication between discursive communities inescapably involves the construction of social relations through the performance of community.

One of the most difficult aspects of the interpretive approach I have chosen is to avoid speaking of the 'characteristic' explanatory behaviour of a discursive community. It becomes tempting to say 'academics say this' or 'policy makers say that', as if these were descriptions (or, worse, defining descriptions) of the behaviour of these groups. In

principle, I assume that it is open to a member of a discursive community to renegotiate the boundaries and terms of membership of a discursive community at any moment; however in practice one would expect to see continuity of explanatory behaviour, and indeed without some repetition and sharing of formulations it might become difficult to speak of a discursive community. Moreover, insofar as my case study is a form of anthropological study, I have been looking for such shared formulations within discursive communities. I have suggested, for example, that members of the AI industry explain AI in terms of market prospects (3.1; 3.2), or in terms of system specifications and target applications (3.2); that venture funders explain AI in terms of financial risk (7.1); that policy makers explain AI in terms of technological trajectories and international competitiveness (7.2; 7.3); and that AI researchers explain AI in terms of machine intelligence (4.1; 6.1; 6.2), which may be contrasted to or assimilated to human intelligence (4.2; 5.1), or in terms of the symbolic paradigm (5.1; 5.2; 7.1), or as an approach to programming (3.1; 3.2; 7.1). These different 'characteristic' explanatory practices are certainly important to my case, since they illustrate the more general fact that explanatory practices are relative to discursive communities and to their social-institutional location. However, my general argument is that discursive communities achieve their identity through a continuing negotiation of 'who we are' and 'what we do'. In this sense, a discursive community has no transcendent identity, at least for the methodological purposes of sociological interpretation. Nonetheless, it is not the case that anything goes in interpreting the explanatory practices of a discursive community: if, as Barbara Herrnstein Smith (1988, p 9)[202] argues, 'Evaluation is always compromised because value is always in motion', nonetheless any specific evaluation or interpretation cannot escape attempting to be accurately drawn or rigorously argued, if it is to convince. That is, for my general point to stand I need to have demonstrated that performative mechanisms are at work in the explanatory practices of the narratives that I have used in my case study; however, the general case only convinces through the accuracy of specific interpretations, and here I would stand by my anthropological insights, at least until convinced by alternative interpretations.[203]

Finally, I address the question of what my study contributes to discussions of

---

[202] These remarks were made in the context of an essay on evaluating Shakespeare's Sonnets (Herrnstein Smith, 1988, pp 1-9). Nonetheless, they seem relevant to more general questions of interpretation, particularly in the light of the STS concern with interpretive flexibility (cf sections 1.2; 6.2).

[203] It is possible to take the 'methodological horrors' (Woolgar 1988a) as implying that any reading is possible of a text and hence that the text 'has' no meaning. Here I am trying to avoid this implication by suggesting that while any particular interpretation can be challenged, this is not tantamount to a challenge to interpretation in general.

technology transfer or communication between discursive communities. In Chapter One I argued that there is a problem in the technology policy literature of describing the local appropriation of knowledge in technology transfer. I suggested that this was reflected in discussions in the STS literature, where the question might be posed as a problem of describing how users' needs are constructed through technological design (cf Latour 1987; Woolgar, 1991), while also allowing for the relative stability of some social groups involved in technology transfer. I suggested that this might be approached through conceptualising technology transfer as communication between discursive communities (cf Woolgar, 1994b), and developing this in terms of the idea of performance of community (cf Cooper and Woolgar, 1993), as well as Schegloff's (1972) description of the way in which selection and recognition of formulations of location may be used to test membership of a community. From this point of view, my central claim for my case study is that I have shown that the explanatory practices associated with AI are a site in which authors may do any of the following (in any combination): adduce who 'we' are (negotiate and maintain community); adduce what AI is, and what it is for; and construe who the audience is and what they need or want to be told about AI (configure the user). In Chapter Three, for example, I showed how the indigenous commentators performed membership through their selection of issues (what counts as a 'story') that addressed the AI industry as audience as well as subject matter. In Chapter Four, I showed both how different narrative registers were relative to the construed audience, and that the identity of the field was not constant for different audiences. In Chapter Five, I showed how how different narratives of the history of AI acted either to compound power, by denying some available representations of difference, or to marginalise subfields by trying to suppress specific readings or groups of readers. In Chapter Six I argued that academic disciplines may be seen to perform community through imposing a preferred reading that marginalises other readings. I also explored some varying ways in which the AI research community handled attributions of 'intelligence' relative to audience (on the whole it figured in explanations for outsiders more than for insiders). While the latter point is specific to the case of AI (and it may not be easy to find equivalent attributes, even for other research areas making claims of radical social impact), the other examples illustrate a generalisable case concerning the performance of community in explanatory practices. This suggests that differences in understanding and explaining research products, relative to a discursive community, are not aberrant cases, but are likely to pervade all communication between different discursive communities.

I also argued that, in the case of AI, there was no neutral or common description which could produce agreement across discursive boundaries. Both the industrial and academic AI communities frequently deployed a technical description of AI (in terms of programming paradigms, for example) to act as a minimal identifying description. For example, Ovum (1989) referred to a technical description to protect AI as knowledge-based systems from the human-like implications of the phrase 'expert systems' (cf 3.1); and the *The Handbook of AI* (Barr and Feigenbaum , 1981; 1982) deployed technical descriptions of AI as a form of programming, in enrolling industrial researchers as colleagues (cf 7.1). While these formulations may be said to have provided a minimal common identity for AI *within* the AI communities, I argued (in 7.1) that they failed to produce a common identity that could be agreed across discursive boundaries. In particular, they produced a construal of the applications of AI (as human like systems) which did not effectively produce agreement across discursive boundaries (for example, it was not necessarily obvious to corporate executives that these were a cost-effective solution for their immediate practical problems). At the same time, however, I suggested that the very futuristic feel of the construed applications, enabled political and industrial strategists to identify slightly less 'human' construals of the future of computing as a measured judgement. Again, some details of this story are not generalisable, and one of the generalisable lessons (the need to create or configure users for an innovatory technology) is already a familiar one (cf Latour, 1987; Woolgar, 1991). What this case study adds, I would argue, is further understanding of the lure of technological determinism for technologists, and how this may act to close off the imagined future. Nelson and Winter (1982, pp 258-259) and Dosi (1984, p 15) take the closing off of the imagined future to be an explanation of relative stability in technological development. However, I have shown that the AI community was taken by surprise by the failure of AI markets, because they supposed that human-like systems must be the future of computing, and that the applications of AI followed directly from the technology. The remark by *The Spang Robinson Report* (Aug 1988, p 2) that 'AI is a technology not a market' (cf 3.2), and by Teknowledge on its 1999 web page that 'while AI is a very powerful technology, it is not an industry in its own right' (cf 3.3), may be read as acknowledging the limitations of a technical description to act as a carrier of a technical future.

To conclude, I would also point out some more general implications for technology

policy of recognising that the performance of community is prevalent in explanatory practices and in communication between discursive communities. This challenges the assumption that differences in understanding, valuing and applying the products of research are in themselves a problem, or involve 'cultural' issues that need to be changed (cf Vaux et al, 1999). My discussion of the case of AI suggests that the local appropriation of knowledge is not an aberration but an element of both successful and unsuccessful communication between discursive communities. This suggests the possibility of more detailed analysis of the lessons of specific case histories. One interesting way to continue from the present case study would be a study of ways in which the Alvey Programme is now evaluated, both as a precedent and as a source of lessons, through interviews with government policy makers (past and present) and members of the AI community.

BIBLIOGRAPHY

Adler, Frederick, 1984. "An Investment Opportunity?" in Winston and Prendergast (1984).

Altorki, Soraya, 1982. "The Anthropologist in the Field: A case of `Indigenous Anthropology' from Saudi Arabia" in Fahim (1982).

Anderson, J R and G H Bower, 1973. *Human Associative Memory.* Washington DC, Winston.

Ashby, W R, 1947. "The Nervous System as Physical Machine with Special Reference to to the Origin of Adaptive Behaviour" in *Mind*, LVI, pp 44-59.

Barr, Avron & Edward A Feigenbaum, 1981. *Handbook of Artificial Intelligence.* Volume One. Reading, Mass., Addison Wesley.

Barr, Avron & Edward A Feigenbaum, 1982. *Handbook of Artificial Intelligence.* Volume Two. Reading, Mass., Addison Wesley.

Bateson, Gregory, 1973. *Steps to an Ecology of Mind.* Paladins, St Albans. (Orig. 1972, Chandler Publishing Co)

Berleur, Jacques, Andrew Clement, Richard Sizer and Diane Whitehouse (eds) 1990. *The Information Society: Evolving Landscapes.* New York, Springer Verlag.

Bieber, Marion, 1985. *Government, University and Industry; Reconciling their Interests in Research and Development.* (Special Report No.214) London, The Economist Intelligence Unit.

Bijker, Wiebe E, Thomas P Hughes and Trevor Pinch (eds) 1987. *The Social Construction of Technological Systems.* Cambrige, Mass, MIT Press.

Bjerknes, Gro, Pelle Ehn and Morten Kyng, 1987. *Computers and Democracy.* Avebury, Gower Publishing Co Ltd.

Block, H David, 1970. "A Review of Perceptrons". *Information and Control*, Vol 17, pp 510-22.

Bloomfield, Brian, ed, 1987. *The Question of Artificial Intelligence.* Beckenham, Croom Helm.

Bloor, David, 1976. *Knowledge and Social Imagery.* Chicago, University of Chicago Press. (Page references to 2nd edition, 1991.)

Boden, Margaret, 1977. *Artificial Intelligence and Natural Man.* Brighton, Harvester Press.

----------, 1988. *Computer Models of Mind.* Cambridge, Cambridge University Press.

Bruner, Jerome, 1983. *In Search of Mind.* New York, Harper & Row.

Bundy, Alan, 1988. "What is the Well-Dressed AI Educator Wearing Now?" in Engelmore, 1988.

Charles, David and Jeremy Howells, 1992. *Technology Transfer in Europe*. London, Belhaven Press.

Callon, Michel, John Law and Arie Rip (eds), 1986. *Mapping the Dynamics of Science and Technology*. London, MacMillan Press.

Callon, Michel, 1986. "The Sociology of an Actor Network: The Case of the Electric Vehicle". In Callon, Law and Rip (1986).

------ 1987. "Society in the Making: The Study of Sociology as a Tool for Sociological analysis". In Bijker, Hughes and Pinch (1987).

------ 1992. "The dynamics of techno-economic networks". In Coombs et al, 1992.

------ 1993. "Variety and irreversibility in networks of technique conception and adoption." In Foray and Freeman, 1993.

Callon, Michel and Bruno Latour, 1992. "Don't throw the baby out with the Bath School! A reply to Collins and Yearley." In Pickering, 1992.

Clancey, William J, Stephen W Smoliar and Mark J Stefik, 1994. *Contemplating Minds*. Cambridge, Mass, MIT Press.

Clark, Peter and Neil Staunton, 1989. *Innovation in Technology and Organisation*. London, Routledge.

Clifford, James and George E Marcus, 1986. *Writing Culture. The Poetics and Politics of Ethnography*. Berkeley and Los Angeles, University of California Press.

Cohen, Paul R & Edward A Feigenbaum, 1982. *Handbook of Artificial Intelligence*. Volume Three. Reading, Mass., Addison Wesley.

Collins, Harry M, 1985. *Changing Order. Replication and Induction in Scientific Practice*. Chicago, Univ of Chicago Press.

----------, 1987a. "Expert Systems and the Science of Knowledge" in Bijker et al, 1987.

----------, 1987b. "Expert Systems, Artificial Intelligence and the Behavioural Co-ordinates of Skill" in Bloomfield, 1987.

----------, 1990. *Artificial Experts*. Cambridge, Mass. and London, MIT Press.

Collins, Harry M, and Steven Yearley, 1992. "Epistemological chicken", in Pickering ed, 1992.

Cooley, Mike, 1987. *Architect or Bee? The Price of Technology*. London Chatto and Windus.

Coombs, Rod, Paolo Saviotti and Vivien Walsh (eds), 1987. *Economics and Technological Change*. London, MacMillan.

---------- (eds), 1992. *Technological Change and Company Strategies*. London, Academic

Press.

Cooper, Geoff and Steve Woolgar, 1993. *Software Quality as Community Performance.* CRICT Discussion paper No 37.

Coulter, J, 1983. *Rethinking Cognitive Theory.* London, MacMillan.

Crapanzo, Vincent, 1986. "Hermes' Dilemma: The Masking of Subversion in Ethnographic Description" in Clifford and Marcus (1986).

Crevier, Daniel, 1993. *AI, the Tumultuous History of the Search for Artificial Intelligence.* New York, Basic Books.

Cyert, R M and J G March, 1963. *A Behavioural Theory of the Firm.* Englewood Cliffs, NJ, Prentice Hall.

Dalyell, Tam, 1983. *A Science Policy for Britain.* Harlow, Longmans.

DARPA, 1983a. *Strategic Computing. New-Generation Computing Technology: A Strategic Plan for its Development and Application to Critical Problems in Defense.* Washington, Defense Advanced Research Projects Agency (28 October 1983).

-------, 1983b. *Strategic Computing. Executive Summary.* Washington, Defense Advanced Research Projects Agency (28 October 1983).

-------, 1985. *Strategic Computing. First Annual Report.* Washington, Defense Advanced Research Projects Agency.

Davis, Randall, 1984. "Amplifying Expertise with Expert Systems". In Winston and Prendergast (1984).

Dennet, Daniel C, 1988. "When Philosophers Encounter Artificial Intelligence" in Graubard (1988)

Department of Industry, 1982. *A Programme for Advanced Information Technology.* The Report of the Alvey Committee, London, HMSO.

Derthick, Mark, 1994. "Review of Pinker and Mehler (eds), Connections and Symbols." in Clancey et al (1994).

Dosi, Giovanni, 1982. "Technical Paradigms and Technological Trajectories - a Suggested Interpretation of the Determinants and Directions of Technical Change". *Research Policy.* Vol 11, no 3.

-------- 1984. *Technical Change and Industrial Transformation.* London, MacMillan.

-------- 1988. "The Nature of the Innovative Process" in Dosi et al (1988).

Dosi, Giovanni, Christopher Freeman, Richard Nelson, Gerald Silverberg and Luc Soete (eds), 1988. *Technical Change and Economic Theory.* London and New York, Pinter Publishers.

Dreyfus, Hubert, 1972, 1979. *What Computers Can't Do.* New York, Harper & Row.

--------- 1992. *What Computers Still Can't Do.* Cambridge, Mass, MIT Press.

Dreyfus, Hubert and Stuart Dreyfus, 1986. *Mind Over Machine.* New York, The Free Press; Oxford, Basil Blackwell.

Elias, Norbert, Herminio Martins and Richard Whitley, 1982. *Scientific Establishments and Hierarchies.* Sociology of the Sciences yearbook 1982, Dordrecht, D Reidel.

Elliott, Brian, ed, 1988. *Technology and Social Process.* Edinburgh, Edinburgh UP.

Engelmore, Robert (ed), 1988. *Readings from the AI Magazine.* Volumes 1-5, 1980-1985. Menlo Park, Ca, AAAI.

Etkowitz, Henry, Andrew Webster and Peter Healey (eds), 1997. *Capitalizing Knowledge: The Growth of Academic-Industry Relations.* Albany, State University of New York Press.

Fahim, Hussein (ed), 1982. *Indigenous Anthropology in Non-Western Countries.* Durham, North Carolina, Carolina Academic Press.

Faulkner, Wendy, 1994. "Conceptualising knowledge used in innovation: a second look at the science-technology distinction and industrial innovation." *Science, Technology, and Human Values.* Vol 19, No 4.

Faulkner, Wendy and Jacqueline Senker, 1995. *Knowledge Frontiers. Public Sector Research and Industrial Innovation in Biotechnology, Engineering Ceramics and Parallel Computing.* Oxford, Clarendon.

Feigenbaum, Edward A, 1992. "A Personal View of Expert Systems: Looking Back and Looking Ahead" in *Expert Systems with Applications* Vol 5 pp 193-201. Oxford, Pergamon Press Ltd.

Feigenbaum, Edward A and Julian Feldman, eds,1963. *Computers and Thought.* New York, McGraw-Hill.

Feigenbaum, Edward A and Pamela McCorduck, 1983 (2nd edn 1984). *The Fifth Generation.* Addison Wesley. (Page references to expanded 2nd edition, 1984, London, Pan Books.)

Fleck, James, 1982. "Development and Establishment in Artificial Intelligence", in Elias et al, 1982.

---------, 1987. "Development and Establishment in Artificial Intelligence" in Bloomfield, 1987.

Foray, Dominique and Christopher Freeman, eds, 1993. *Technology and the Wealth of Nations.* London, Pinter Publishers.

Freeman, C, 1982 (1st edn 1974). *The Economics of Industrial Innovation.* London, Frances Pinter.

-------, ed, 1984. *Long Waves in the World Economy.* London, Frances Pinter.

-------, 1988a. "Induced Innovation, Diffusion of Innovations and Business Cycles" in Elliott (1988).

-------, 1988b. Preface to Dosi et al, 1988.

Freeman, C and C Perez, 1988. "Structural crises of adjustment: business cycles and investment behaviour" in Dosi et al, 1988.

Freeman, C and L Soete, 1990. "Information Technology and the Global Economy" in Berleur et al (1990).

Friedman, Milton, ed, 1953. *Essays in Positive Economics*. University of Chicago Press, Chicago.

Fuchi, Kazuhiro and Kunio Murakami, 1983. "Japan's Fifth Generation Computer Project: Progress Report from ICOT" in SPL-Insight (1983).

Gannon, Thomas F, 1983. "Background Paper on the Micro Electronics and Computer Technology Corporation (MCC)" in SPL-Insight (1983).

Gardner, Howard, 1987. *The Mind's New Science.* Basic Books.

Garfinkel, Harold and Harvey Sacks, 1970. "On Formal Structures of Practical Actions", in McKinney and Tiryakian (1970).

Geertz, Clifford, 1983. *Local Knowledge.* New York, Basic Books. (Page references to 1993 edition; London, Fontana Press.)

George, Frank, 1979. *Man the Machine.* Paladin, St Albans, Herts.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott and Martin Trow, 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.* London, Sage.

Gilbert, G Nigel and Michael Mulkay, 1984. *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse.* Cambridge, Canbridge University Press.

Gill, Karamjit (ed), 1986. *Artificial Intelligence for Society.* Chichester, John Wiley.

Gîranzon, Bo and Ingela Josefson (eds), 1988. *Knowledge, Skill and Artificial Intelligence.* London, Springer-Verlag.

Graubard, Stephen R, ed, 1988. *The Artificial Intelligence Debate. False Starts, Real Foundations.* MIT Press, Cambridge, Mass. (Originally published as Daedalus Vol 117, No 1, Winter 1988, The American Academy of Arts and Sciences.)

Grint, Keith and Steve Woolgar, 1997. *The Machine at Work.* Cambridge, Polity Press.

Guy, Ken and Luke Georghiou, 1991. *Evaluation of the Alvey Programme for Advanced Information Technology.* London, HMSO.

Hallam, John and Chris Mellish, 1987. *Advances in Artificial Intelligence.* Chichester, John

Wiley.

Heims, Steve J, 1991. *The Cybernetics Group.* MIT Press, Cambridge, Mass.

Henkel, Mary, Stephen Hanney, Maurice Kogan, Janet Vaux and Dagmar von Walden Laing, 1999. *Academic Responses to the UK Foresight Programme.* Report presented to the Nuffield Institute, London.

Herrnstein Smith, Barbara, 1988. *Contingencies of Value. Alternative Perspectives for Critical Theory.* Harvard University Press, Cambridge, Mass and London.

Hirooka, M, 1986. "The Role and Development of the Chemical Industry in the Innovation Age", paper presented at Third World Congress of the Society of Chemical Engineering, Tokyo (cited in Martin and Irvine, 1989)

Hodges, Andrew, 1983. *Alan Turing. The Enigma of Intelligence.* Burnett Books Ltd, London. (Page references to 1985 edn; London, Unwin Paperbacks.)

Hofstadter, Douglas R, 1979. *Gödel, Esher, Bach: An Eternal Golden Braid.* New York, Basic Books.

Hook, Sidney, ed, 1960. *Dimensions of Mind.* New York, New York University Press.

House of Lords, 1985. *ESPRIT (European Strategic programme for Reserch and Development in Information Technology).* Report by Select Committee on the European Communities. 8th Report, Session 1984-85. London, HMSO.

Hughes, Thomas P, 1987. "The Evolution of Large Technological Systems" in Bijker, Hughes and Pinch (1987).

Innis, Harold, 1951. *The Bias of Communication.* Toronto, University of Toronto Press.

Janeway, William, 1984. "Financing the Future" in Winston and Prendergast (1984).

Kleene, Stephen Cole, 1952. *Introduction to Metamathematics.* New York, Van Nostrand.

Kuhn, Thomas S, 1962 (2nd edn,1970).*The Structure of Scientific Revolutions.* Chicago, University of Chicago Press.

------- 1970. "Postscript - 1969" in 2nd edn, *The Structure of Scientific Revolutions.* Chicago, University of Chicago Press.

Latour, Bruno, and Steve Woolgar, 1979 (2nd edn. 1986) *Laboratory Life. The Construction of Scientific Facts.* Princeton UP, Princeton.

Latour, Bruno, 1987. *Science in Action.* Harvard UP, Cambridge, Mass.

------- 1988 (a). *The Pasteurization of France.* (Trans Alan Sheridan and John Law.) Harvard UP, Cambridge, Mass. and London.

------- 1988 (b). "The Prince for Machines as well as for Machinations". In Elliot (1988)

------- 1991 (English translation 1993). *We Have Never Been Modern.* Hemel Hempstead,

Harvester Wheatsheaf.

Law, John, 1987. "Technology and Heterogeneous Engineering: the Case of Portuguese Expansion". In Bijker, Hughes and Pinch, 1987.

----- (ed), 1991. *A Sociology of Monster. Essays on Power, Technology and Domination.* London and New York, Routledge.

Leydesdorff, Loet and Henry Etkowitz, 1996. "The Future Location of Research: A Triple Helix of University-Industry-Government Relations II" (Theme paper for a conference in New York City, January 1998), EASST Review, Vol 15.4, Dec 1996, pp20-25.

Lighthill, James, 1973. "Artificial Intelligence. A General Survey" in *Artificial Intelligence: A Paper Symposium.* London, Science Research Council.

Lynch, Michael, 1993. *Scientific Practice and Ordinary Action.* Cambridge, CUP.

Lynch, Michael and Steve Woolgar (eds) 1991. *Representation in Scientific Practice.* Cambridge, Mass and London, MIT Press.

MacKenzie, Donald and Judy Wajcman (eds), 1985. *The Social Shaping of Technology.* Milton Keynes, Open University.

MacKenzie, Donald, 1990 (pb 1993). *Inventing Accuracy. A Historical Sociology of Nuclear Missile Guidance.* Cambridge, Mass and London, MIT Press.

-------- 1992. "Economic and sociological explanation of technical change." In Coombs et al 1992.

-------- 1996. *Knowing Machines.* Essays on Technical Change. Cambridge, Mass., MIT Press.

McCarthy, John, 1959. "Programs with Common Sense" in NPL 1959, pp 75-91.

--------- 1978. "History of Lisp". History of Programming Languages Conference. ACM SIGPLAN Notices, Vol 13, No 8. August 1978. (215-223)

---------- 1988 "Mathematical Logic in Artificial Intelligence" in Grauber (1988).

McCarthy, John, Marvin L Minsky, Nathaniel Rochester and Claud E Shannon, 1955. *Proposal for Dartmouth Summer Research Project on Artificial Intelligence.* Proposal to the Rockefeller Foundation, 31 August 1955. (www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html)

McCarthy, J and P Hayes, 1969. "Some Philosophical Problems from the Standpoint of Artificial Intelligence" in B Meltzer and D Michie (1969)

McCorduck, Pamela, 1979. *Machines Who Think.* San Francisco, W H Freeman & Co.

McKinney, John C and Edward A Tiryakian, 1970. *Theoretical Sociology.* New York, Appleton-Century-Crofts.

Mandler, G, 1981. "What is Cognitive Psychology? What isn't?" Address to the APA Division of Philosophical Psychology, Los Angeles.

Martin, Ben R and John Irvine, 1989. *Research Foresight. Priority-Setting in Science.* London and New York, Pinter Publishers.

Mays, W, 1952. "Can Machines Think?" *Philosophy.* 27, pp 148-162.

Meltzer, B and D Michie (eds) 1969. *Machine Intelligence.* Vol 4. Edinburgh UP.

Merton, Robert, 1957. "Priorities in Scientific Research". *American Sociological Review.* 22, 6. (Dec).

Merleau-Ponty, Maurice, 1945. *Phenomenologie de la perception.* (Eng trs, Colin Smith, 1962. *Phenomenology of Perception*, London, Routledge and Kegan Paul.)

Merton, Robert, 1973 (ed Norman W Storer). *The Sociology of Science. Theoretical and Empirical Investigations.* Chicago and London, Chicago UP.

Michie, Donald and Rory Johnston, 1984. *The Creative Compute.*, New York, Viking.

Miles, Ian and Kevin Robins, 1992. "Making Sense of Information", in Robins (1992).

Miller, George A, 1956. "The Magical Number Seven, Plus or Minus Two: Some limits on our capacity for Processing Information". *Psychological Review* 63, pp 81-97.

--------, 1979. "A Very Personal History". Talk to Cognitive Science Workshop, MIT.

Minsky, Marvin L, 1956. *Heuristic Aspects of the Artificial Intelligence Problem.* Group Report 34-55, MIT Lincoln Laboratory.

-------, 1959. "Some Methods of Artificial Intelligence and Heuristic Programming" in NPL 1959, pp 3-36.

-------, 1961. "Steps Towards Artificial Intelligence". Proceedings of the IRE, January 1961.

-------, ed,1968. *Semantic Information Processing.* Cambridge, Mass, MIT Press.

-------, 1985. (1988) *The Society of Mind.* New York, Simon and Schuster.

-------, 1988. "Early MIT Artificial Intelligence Memos: An Introduction to the COMTEX Microfiche Edition" in Engelmore 1988.

-------, 1993. Article 14265 of newsgroup comp.ai.philosophy.

-------, 1994a. "Review of Allen Newell, Unified Theories of Cognition" in Clancey et al (1994), 97-108.

-------, 1994b. "Society of Mind: a response to four reviews" in Clancey et al (1994), 308-333.

Minsky, Marvin L and Seymour A Papert, 1988 (1969). *Perceptrons.* Cambridge, Mass, The MIT Press.

Moto-oka, Tohru and Masaru Kitsuregawa, 1984. *The Fifth Generation Computer. The*

Japanese Challenge. Tokyo, Iwanami Shoten. (English translation by FDR Apps; Chichester, John Wiley, 1985).

Mowery, David C and Nathan Rosenberg, 1989. *Technology and the Pursuit of Economic Growth.* Cambridge, Cambridge University Press.

Mulkay, Michael, 1991. *Sociology of Science. A Sociological Pilgrimage.* Buckingham, Open University Press.

Mulkay, Michael and Nigel Gilbert, 1981. "Putting Philosophy to Work" in Philosophy of the Social Sciences (reprinted in Mulkay 1991).

Nelson, Richard R and Sidney G Winter, 1977. "In search of a useful theory of innovation". Research Policy 6(1), 36-77.

--------, 1982. *An Evolutionary Theory of Economic Change.* Cambridge, Mass, Harvard UP.

Nelson, Richard R, 1987. *Understanding Technical Change as an Evolutionary Process.* Amsterdam, Elsevier.

Newell, Allen, 1980. "Physical Symbol Systems". *Cognitive Science* 4, pp 135-183.

--------, 1990. *Unified Theories of Cognition.* Cambridge, Mass, Harvard UP.

Newell, Allen and Herbert A Simon, 1972. *Human Problem Solving.* Englewood Cliffs, Prentice-Hall.

NPL, 1959. *Mechanisation of Thought Processes.* Vols I and II. Proceedings of National Physical Laboratory Symposium No 10, 24-27 November 1958. London, HMSO

Oakley, Brian, 1984. "Alvey and the Academics". *THES*, 22 June 1984.

Oakley, Brian and Kenneth Owen, 1989. *Alvey: Britain's Strategic Computing Initiative.* Cambridge, Mass and London, MIT Press.

Okely, Judith and Helen Callway, 1992. *Anthropology & Autobiography.* ASA Monographs, 29. London, Routledge.

Olazaran, Mikel, 1996. "Official History of the Perceptrons Controversy". *Social Studies of Science,* 26.3, pp 611-659.

OST, 1993. *Realising our Potential - a Strategy for Science, Engineering and Technology.* Command 2250. London, HMSO.

Ovum, 1984. *The Commercial Application of Expert Systems* (by Tim Johnson). London, Ovum.

--------, 1986a. *Expert Systems 1986* Vol 1: USA and Canada (by Julian Hewett and Ron Sasson). London, Ovum.

--------, 1986b. *Commercial Expert Systems in Europe* (by Julian Hewett, Stephen Timms and Geoffroy d'Aumale). London, Ovum.

-------, 1988a. *Expert Systems in Banking and Securities* (by Christine Guilfoyle and Judith Jeffcoate). London, Ovum.

-------, 1988b. *Expert Systems Markets and Suppliers* (by Tim Johnson, Julian Hewett, Christine Guilfoyle and Judith Jeffcoate). London, Ovum.

--------, 1989. *Knowledge-Based Systems: Markets, Suppliers and Products* (by Tim Johnson, Julian Hewett, Christine Guilfoyle, Judith Jeffcoate and Candice Goodwin). London, Ovum.

Papert, Seymour, 1988. "One AI or Many?" in Graubard (1988).

Penrose, Roger, 1989. *The Emperor's New Mind.* Oxford, Oxford University Press. (Page references to 1990 edition, London, Vintage.)

Perez, Carlotta, 1983. "Structural Change and the Assimilation of New Technologies in the Economic and Social Systems". *Futures,* October 1983, pp 357-375.

-------- 1985. "Microelectronics, long waves, and world structural change". *World Development.* 13.3, pp 441-63.

Petrella, Riccardo, 1990. "Growth, Productivity and Innovation. Theories and Facts" in Berleur et al (1990).

Pickering, Andrew, ed, 1992. *Science as Practice and Culture.* Chicago and London, Chicago UP.

Pinch, Trevor J and Wiebe E Bijker, 1989. "The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other", in Bijker et al 1989.

Polanyi, M, 1967. *The Tacit Dimension.* London, Routledge and Kegan Paul.

Posner, M and G L Shulman, 1979. "Cognitive Science" in E Hearst, ed, *The First Century of Experimental Psychology.* Hillside, NJ, Lawrence Erlbaum.

Putnam, Hilary, 1959. "Minds and Machines". In Hook (1960).

----------, 1988a. *Representation and Reality.* Cambridge, Mass. MIT Press.

----------, 1988b. "Much Ado About Not Very Much" in Graubard (1988).

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik, 1985. *A Comprehensive Grammar of the English Language.* London, Longman

Robins, Kevin (ed), 1992. *Understanding Information.* Business, Technology and Geography. London and New York, Belhaven Press.

Rorty, Richard, 1967. *The Linguistic Turn: Recent Essays in Philosophical Method.* Chicago, University of Chicago Press.

Rosaldo, Renato, 1986. "From the Door of His Tent: The Fieldworker and the Inquisitor" in

Clifford and Marcus (1986).

Rosenberg, Nathan, 1982. *Inside the Black Box: Technology and Economics.* Cambridge, Cambridge UP.

-------- 1994. *Exploring the Black Box. Technology, Economics and History.* Cambridge, Cambridge UP.

Rosenblatt, Frank, 1959. "Two Theorems of Statistical Separability in the Perceptron" in *Mechanisation of Thought Processes* (NPL, 1959)

Rosenbrock, Howard (ed), 1989. *Designing Human-centred Technology.* London, Springer-Verlag.

Rothwell, Roy and Walter Zegveld, 1981. *Industrial Innovation and Public Policy.* London, Frances Pinter Ltd.

Sacks, Harvey, 1972. "An Initial Investigation of the Usability of Conversational Data for Doing Sociology", in Sudnow (1972)

Scarbrough, Harry and J Martin Corbett, 1992. *Technology and Organisation. Power, Meaning and Design.* London, Routledge.

Schaffer, Simon, 1998. *Babbage's Intelligence: Calculating Engines and the Factory System.* http://www.wmin.ac.uk/media/schaffer

Schegloff, Emanuel A, 1972. "Notes on a Conversational Practice: Formulating Place". In Sudnow (1972)

----------, 1997. "Whose text? Whose context?" *Discourse and Society.* Vol 8, No 2, pp 165-187.

Scherer, F M, 1984. *Innovation and Growth. Schumpeterian Perspectives.* Cambridge, Mass, MIT Press.

Schmoch, U, S Hinze, G Jackei, N Kirsch, F Mayer-Crammer and G Munt, 1993. *Constraints and opportunities for the dissemination and exploitation of R&D activities: the R&D environment. Structural features and their modelling.* Karlsruhe, Fraunhofer Institute.

Schumpeter, Joseph, 1942. *Capitalism, Socialism, and Democracy.* New York, Harper & Row.

Schwartz Cohen, Ruth, 1985. "How the refrigerator got its hum", in MacKenzie and Wajcman (1985)

Science and Engineering Research Council and Department of Industry, 1983. *Intelligent Knowledge Based Systems. A Programme for Action in the UK.* Vol I. Report of the IKBS Architecture Study. Swindon, SERC.

Searle, John, 1980. "Minds, Brains and Programs", *Behavioural and Brain Sciences* 3, 417-

458.

----------, 1984. *Minds, Brains and Science.* The 1984 Reith Lectures. London, BBC.

Segal, Quince & Partners, 1985. *The Cambridge Phenomenon.* Cambridge, Segal Quince & Partners.

Selfridge, Oliver, 1959. "Pandemonium: a Paradigm for Learning" in NPL 1959, pp 511-531.

Shanker, S G, 1987. "AI at the Crossroads" in Bloomfield (1987).

Shapin, Steven and Simon Schaffer, 1985. *Leviathan and the Air-Pump: Hobbes, Boyle and the Experimental Life.* Princeton, NJ, Princeton UP.

Simon H A, 1947. (1956,1976). *Administrative Behaviour.* New York, Macmillan.

------ 1969 (2nd Edition 1981). *The Sciences of the Artificial.* Cambridge, Mass., The MIT Press.

------ 1991. *Models of My Life.* New York, Basic Books.

Sluckin, Wladyslaw, 1954 (2nd edition 1960). *Minds and Machines.* Penguin Ltd, Harmondsworth.

Sokal, Alan, 1996. "Transgressing the Boundaries: Toward a transformative hermeneutics of of quantum gravity". *Social Text,* 46/47, pp 217-252.

Sokal, Alan and Jean Bricmont, 1998. *Intellectual Impostures.* London, Profile Books.

SPL-Insight, 1983. *Fifth Generation World Conference.* Proceedings of conference held in London, UK, 27-29 September 1983.

SPL-Insight, 1984. Impact 1984. *New Horizons from Fifth Generation Computing.* Proceedings of conference held in London, UK, 17-18 September, 1984.

SRI International, 1984. *Advanced Information Technology.* Proceedings of conference held in Churchil College, Cambridge, UK, 20-22 September 1984.

Stankiewicz, Rikard, 1986. *Academics and Entrepreneurs: Developing University-Industry Relations.* London, Frances Pinter.

Stewart, Alex, 1985. *Automating Intelligence. The Fifth Generation.* Japan Focus. London, Baring Far East Securities Ltd

Stokes, Adrian V, 1986. *Concise Encyclopedia of Information Technology* (3rd edn). Aldershot, Wildwood House.

Sudnow, David (ed), 1972. *Studies in Social Interaction.* The Free Press, New York; Macmillan, London.

Takei, Kinji, 1984. "Progress in the Initial Stages of the FGCS Project". In SRI 1984.

Teichman, Jenny, 1988. *Philosophy and the Mind.* Oxford, Basil Blackwell.

Tisdell, C A, 1981. *Science and Technology Policy: Priorities of Governments.* London and

New York, Chapman and Hall.

Tsatsaroni, Anna and Geoff Cooper, 1999. "Transformation as Knowledge: Deconstruction, SSK and Science Education". Paper presented to conference on The Transformation of Knowledge, University of Surrey, January 12-13, 1999.

Turing, Alan, 1937. "On Computable Numbers, with an application to the Entscheidungsproblem". *Proc. London Math. Soc.* (2) 42, 230-265.

---------, 1950. "Computing Machinery and Intelligence". *Mind* LIX.236, October 1950.

Turkle, Sherry, 1984. *The Second Self. Computers and the Human Spirit.* London and New York, Granada.

Van den Belt, Henk and Arie Rip, 1989. "The Nelson-Winter-Dosi Model and Synthetic Dye Chemistry" in Bikjer, Hughes and Pinch (1987).

Vaux, Janet, 1986. "AI and Philosophy: Recreating Naive Epistemology" in Gill (1986).

Vaux, Janet, and Paula Gomes, Jean-Noel Ezingeard, Robert Grieve and Steve Woolgar, 1999. "Managing to avoid innovation. Problems of technology transfer in small firms". *Industry and Higher Education.* Vol 13.1, pp 25-32.

Vergragt, Philip J, 1988. "The social shaping of industrial innovations". *Social Studies of Science.* Vol 18, 483-513.

Vergragt, Philip J, Peter Groenwegen and Karel F Mulder, 1992. "Industrial Technological Innovation" in Coombs et al (1992).

Von Neumann, John, 1958. *The Computer and the Brain.* New Haven. Yale University Press.

Webster, Andrew and Henry Etzkowitz, 1991. *Academic-Industry Relations: The Second Academic Revolution?* SPSG Concept Paper 12. London, Science Policy Support Group.

Weizenbaum, Joseph, 1976. *Computer Power and Human Reason.* W H Freeman & C0. (1984, with new preface, Harmondswork, Penguin Books.)

Weiner, Norbert,1948, 1961. *Cybernetics.* MIT Press, Harvard, Mass.

-------- 1950, 1968. *The Human Use of Human Beings.* Houghton Mifflin Company (Page refs to 1968 edn, Sphere Books, London)

White, Lynn, 1962. *Medieval Technology and Social Change.* Oxford, Oxford University Press.

Winograd, Terry, 1972. *Understanding Natural Language.* New York, Academic Press.

Winograd, Terry and Fernando Flores, 1986. *Understanding Computers and Cognition.* Norwood, New Jersey, Ablex Publishing Corporation.

Winston, Patrick H and Karen A Prendergast, 1984. *The AI Business. The Commercial Uses*

of Artificial Intelligence. Cambridge, Mass and London. MIT Press.

Woolgar, Steve, 1985. "Why not a sociology of machines? The case of sociology and artificial intelligence". *Sociology* 19.4, 557-572.

------- 1988a. *Science: The Very Idea.* Chichester and London, Ellis Horwood and Tavistock.

------- (ed) 1988b. *Knowledge and Reflexivity: New Frontiers in the Sociology of knowledge.* London. Sage.

------- 1989. "Reconstructing man and machine: a note on sociological critiques of cognitivism" in Bijker, Hughes and Pinch (1989)

------- 1991. "Configuring the user: the case of usability trials" in Law (1991).

------ 1994a. *Science and Technology Studies and the Renewal of Social Theory.* CRICT Discussion Paper No 41.

------ 1994b. *Rethinking the Dissemination of Science and Technology.* CRICT Discussion Paper No 44

Yearley, Steven, 1988. *Science, Technology and Social Change.* Boston, Unwin Hyman

OTHER PUBLICATIONS

Periodicals (AI newsletters)

AI Business. London, Compass Press; and Tonbridge, Kent, Industrial Media Ltd.

AI Watch. Oxford, AI Intelligence.

Expert Systems User. London, Compass Press; and Tonbridge, Kent, Industrial Media Ltd.

Expert Systems. Oxford, Learned Information.

Intelligent Software Strategies. (Formerly, *Expert Systems Strategies*). Arlington. MA, Cutter Information Corp.

Intelligent Systems Report. (Formerly *ICS Applied Aritifical Intelligence Reporter*, and *AI Week*). Miami, University of Miami; and Atlanta, GA, AI Week Inc.

La Lettre de L'Intelligence Artificielle. Paris, SARL and EC2.

Machine Intelligence News. (Now incorporated in *AI Watch,* qv) London, IBC; and JV Publications.

Spang Robinson Report on Intelligent Systems. Palo Alto, Spang Robinson; and New York, John Wiley.