

## AI and Representation: A Study of a Rhetorical Context for Legitimacy

Lynette A. C. Hunter

*School of English, University of Leeds, Leeds, UK*

**Abstract:** Theoretical commentaries on AI often operate as a metadiscourse on the way which science represents itself to a wider public. The sciences and humanities do the same of work but in different fields that encourage them to talk about their work differently: science refers to a natural world that does not talk back, and the humanities refer continually to a world with communicative people in it. This paper suggests that much AI commentary is misconceived because it models itself on the way that science represents itself, rather than on the way that the humanities address social context.

AI theorists have become increasingly worried about the lack of evaluation in AI, the lack of reflexivity, and the lack of contact with society. Frequently these writers turn to concepts of tacit knowledge to work through these worries. In doing so they are recognising the problem of AI's second-order representation of science and trying to deal with it. However, this recognition of a problem with the representations of science simply turns back to the legitimization crisis in Western politics where many commentators use science precisely as a 'model' for other political institutions. They do so because science is one of the few areas of knowledge which has been legitimate to use plausible methodology for representation that allows for arbitrary designations of authority as well as parallel systems of different authority. However, plausible rejects any control on reflexivity, assumes an ethnocentric club culture and does not address social context.

It is in this sense that the problems of legitimization in political liberalism are similar to those of legitimization in sciences, both are rooted in their uses of representation. AI's link with representation of science places it in the heart of this debate about legitimacy. This paper suggests that AI does need to learn about reflexivity and that it might well do so by looking at the recent work on experimentation and representation by historians of science, and by looking to the debates about representation by historians of science, and by looking to the debates about representation within the humanities. However, reflexivity may not be enough. Devising a way of thumb for the appropriate halting of reflexivity, is also needed to address social context and take action.

**Keywords:** Artificial intelligence; Ethnocentricity; Legitimation; Rhetoric; Science (epistemology of); Social context

# 1. Introduction

In the short history of AI it is evident that it has begun to explore a variety of techniques and methods, but it has yet to learn how to involve reflexivity in order to provide contexts for them. And although reflexivity is a step along the way, the well-considered methodology and processes of assessing social value, in rhetorical terms, AI needs not only an understanding of ethos and pathos, the participant and the responsive context, but an awareness of stance which provides guidelines for handling continual reflexivity in order to make decisions about necessary social action.

In 1991 a lucid and empowered essay by David Kirsh outlined a variety of approaches to AI, and took on the big issues of cognition, representation and learning. In the process of doing so he positions as a central concern, questions to do with how we come to agree to common grounds for representation, and charts the move by some AI researchers from axiom to plural perspectives. While the critique of the logicist assumptions is welcome and in these hands sophisticated, the plurality of results does not address issues of social immediacy and action.

More and more commentators recognise the need in their talk about evaluation: how do we know when AI can be said to be valuable? More to the point, how do we know that value is not just success in fulfilling the internally constituted rules of self-defining worlds, but is rather an intersection with social needs that offers a way of addressing them. Evaluation is pertinent to all areas of inquiry, but particularly so in AI at the moment, because there are so few rules of thumb for the way in which AI deals with representation and touches upon reality.

Possibly because of its intimate relationship with modern science and its attention to stress on method, AI has concentrated on finding techniques and plausible common grounds for its work. In rhetoric plausible argument works from opinion, and argues from basic assumptions held by a group of people. In contrast, probable argument first attempts to discuss commonly held ground, decides upon its appropriate basis for the argument in question, and only then proceeds to argue from it. There has been a lot of commentary on the advantages and disadvantages of plausible and probable argument, with most writers agreeing that the plausible is the root of deception and manipulation. There is also a line that rejects rhetoric entirely and claims that since truth about certain things is absolute and can be known and articulated, basic assumptions do not need to be discussed; in other words the 'plausible' is not relevant to 'certain' things, defined, of course, by the individual and/or hegemonic group on behalf of others. Rhetoricians frequently concur that probable argument is the more testing and helpful, but that plausible argument is often necessary because of time constraints or the need for immediate action.

At least since Aristotle, science and more generally, ordered knowledge, has been described as an activity which need not engage in probable argument all the time. One of the reasons given for this is that within a field of knowledge, people who are part of that field by definition have worked through the basic assumptions. They are pursuing that science only because they have entered a club, a culture or a ground to which they have given prior assent. The problem, of course, is that the activity of assent is soon forgotten for a variety of reasons (Gooding, 1989, p. 2) and the door is closed on the reassessments enabled by probable argument. Indeed, science has come to be known in broad terms as the field of intellectual pursuit.

which people behave as if they are working with 'certainties', or at the very least if there are basic assumptions that do not need to be questioned, that are independent of context and appropriateness.

Kirsh defines AI as based on the description of a consensual reality with adequate vocabulary for articulating concepts that underpin "the millions of things we know and that we assume everyone else knows"<sup>1</sup>. But having been articulated in this way, the description is rarely questioned. It is perhaps significant that the fields of inquiry recognised as impinging most directly on AI, from philosophy and linguistic theory to computer science and cognitive science, are not those demanding scientific immediacy. On the other hand inquiry in the humanities is seen to have retained an abiding interest in precisely what makes for appropriateness and attention to scientific context.

If AI is to assess appropriateness, it has a specific problem because it is perceived as more relevant to the sciences than the arts and humanities. Assessment of the common grounds for representation will have to push open the door to rhetoric that Aristotle pushed to, and will have to move the public debates of science from the plausible to the probable. The plausible can be accepted and even fun, for words of games that do not impinge on social relations of dominance and power. Where the social systems of scientific games become increasingly influential through various technologies and engineering, politics makes assessment essential. The first stage of assessment is to question, both what is appropriate to say and what is necessary to say but difficult.

## 2. Rhetoric, Reflexivity and Stance

A rhetorical approach to the politics of representation that results when the arts and humanities come to deal with computers and computing (Gardin, 1990), sets up a discussion about the illusions of power offered by closed systems, and the problem of knowing when they have ceased to be helpful or whether it is even possible to open them up. This discussion is analogous to the much broader debate about the political implications of communication that has been going on in Western European political philosophy for some time, as it has attempted to cope with the effects of sudden authorization in the early 20th century of new voices from the recently enfranchised population. Suddenly the whole process of how one finds and assesses common grounds, which had been worked out even in the post-medieval world through terms of effective power politics for a small elite, was placed in the world of a widely varied and huge public.

The problem of differentiating between the necessary and immediate agreement about appropriate common grounds (the probable), and habitual agreement about them (the plausible), has been problematised by the sheer size and number of many divergent and conflicting needs and desires. Both consensus argumentation directed to a decision about action in which all the discussants participate (Gadamer, 1975) and corporate argumentation in favour of an action which will be taken on behalf of the participants (Habermas, 1987), may lead to tacit agreement about adequate representation. The latter shift from corporate to totalitarian argument has been well-documented,<sup>2</sup> but the elision from consensus to the self-enclosed cynicism of

nostalgia of the pluralist worlds of hyperliberal club culture has only recently been described.<sup>3</sup>

As this paper will suggest, the activity of modern science has much in common with club culture but is also working, significantly, in a representative mode that is completely different to that of the humanities. Furthermore, it is not surprising that science has become a central motif in the discussions about 'legitimation crisis' of the state, given that science is the one field where working with plausible argument on opinion, as if it were certain, is expected and authorised by the representative mode. Current descriptions refer to science working within the self-enclosed 'worlds' of 'paradigms', either large-scale Kuhnian structures or the competing pluralities of post-modernism described by Lyotard (1979). If writers such as Gadamer and Habermas proposed, in the 1970s and 1980s respectively, a return to probable argument through assessed consensus discussion, or a return to corporate argumentative universals (Habermas, 1987, p. 194), others, such as Rorty following an implicit line in Althusser, argue that the distinction between the probable and the plausible is not relevant because we all do live in enclosed club cultures, therefore the probable is always opinion: there is no need to distinguish between an appropriate and an inappropriate ground, there is no need to assess social context.

My position here is this: that it is necessary to assess social context, to discuss appropriateness, to attempt to distinguish the probable from the plausible. This paper will argue that AI has been caught, for specific reasons, into duplicating club culture when it could be precisely the place where science attempts a different representative mode (Geiss and Viswanathan, 1986, p. xxi; Woolgar, 1991).

## **Reflexivity and How to Stop it**

If AI is to take a stance on the activities of science and technology that it appears to describe, it needs first to become self-conscious about its own techniques and strategies. AI is about finding ways to represent the activities of human intelligence both how we represent the world to ourselves (knowledge representation) and how we deal with that representation, respond to it, interact with it (expert systems, etc.) and so on. Representation is a complex area, and it is not the intention here to limit it to even sophisticated versions of visual verisimilitude or structural congruence. Rather, I would like to place AI within the carefully detailed debates to which the long history of representation in the arts and humanities has led. This is to understand the rhetorical work of representing which involves not only *technique* and *description* but discussion of *strategy*, and most importantly an assessment of *stance*. It is within this stance that the dilemma about evaluation in humanities computing can be addressed. And it is questions about evaluation in AI that urgently need to be answered because it can sensibly position itself with regard to technology and power. This discussion will focus on issues of self-consciousness, strategy and reflexivity.

A consideration of strategy has to include a sense of where the speaker or author or writer is placed within the description. In traditional rhetoric, 'ethos' implies taking up a particular role or perspective with relation to the audience, whose response was in turn cast as 'pathos'. Ethos implies a fairly stable set of conventions: the relation between the speaker and the discussion; for example, written genre like the letter or report, assume certain strategic relations once they are realised; v

not usually expect a report to be a lyric poem. Generally, within rhetorical studies there is growing contemporary concern with an increasingly varied audience<sup>4</sup> subtly elaborated questions about the effect of the speaker on the discussor. In literary 'deconstruction' and the 'reflexivity' of the social sciences, there has been an extension of ethos to include an on-going assessment of how participation affects effects the event.<sup>5</sup>

If we look reflexively or self-consciously at how our participation in both event and its representation affects that event or representation,<sup>6</sup> questions are readily obviously raised about various parts of our knowledge and perception that are inadequately articulate. From difficult questions, such as 'why do you feel depressed' to apparently simple ones such as 'why do you like cheese?', we often find words and other media inadequate to the representation of what we feel or know. Reflexivity is on-going and continuous precisely because that knowledge cannot satisfactorily be communicated, and this indicates a further problem: reflexivity is a useful tool but its continuous activity does not encourage positions to be taken from which decisions can be made and specific actions carried out.

The problem is how and where to halt reflexivity. On-going questioning is frequently halted at points where articulation appears inadequate. We stop and attend to some kind of communicative context for a practice that resists discussion. But if no attempt is made merely in response to a resistance, then reflexivity simply stalls again and moves on to the next (arbitrary) difficulty. Used in this manner, the process becomes intensely private. It works within a system of absolute/arbitrary dichotomies which define much of the social action that can be taken by the individual as private and science is an intensely private activity.

Articulation and its pair, 'tacit' knowledge, lies at the heart of the debates about legitimation in science: Is tacit knowledge unarticulated because it is impossible to articulate, because people have not yet tried to articulate it, because it has become a habit, because it is blind prejudice, or because it is 'true' and does not need articulation? This range of questions sits parallel with the range of AI approaches listed by Kuhn (1991), from distributed AI to moboticists to connectionists to Soar to logic programming. Without social context it is impossible to distinguish. With private use of reflexivity there is no sense in assessing which of the many difficulties it is necessary to deal with, nor is there any sense of placing an attempt at articulation before a public, for the purposes of evaluation (how appropriate is it in addressing social needs?) or criticism (is it appropriate at all?). This is why strategy, whether stable ethos or on-going reflexivity and deconstruction, needs a stance that places it in a social context.

This version of the 'legitimation crisis' which stresses the problem of distinguishing the habitual from the appropriate public and social common grounds for representation is related to the problem in cognitivist rational analytics of the distinction between habit and prejudice, which is accepted blindly without question, and the axiom or explicitly stated rule (Adorno, 1982). Yet the two versions are not entirely congruent because the cognitivist dilemma does not deal with the difference or distinction between regulative and constitutive rules, rule-following and rule-performing, that makes the debate about habitual/appropriate analogous to the further problem of how to distinguish self-evident from working knowledge (Feyerabend, 1975) or tacit knowledge (Polanyi, 1967), and indeed how to distinguish these two, habitual and appropriate, from each other. These issues are directly relevant to theorising about

evaluation in AI, which has moved inexorably toward concepts of practical knowledge and performance. This has happened largely because the aims of AI to represent knowledge and to represent expertise, cannot be satisfied by equating knowledge to information, or expertise to regulative rule-following. But in order to understand the necessary concepts, AI must learn how to become self-conscious about techniques and methods.

Evaluation for the arts and humanities in the post-medieval period has largely focused on reflexive questioning of common grounds or topics as they are conveyed by representation, even though much humanistic activity does not do this. Philosophy and what is temporarily called theory, has become the recognised place where we attempt to find new common grounds and try out their appropriateness, as well as where we assess the continued appropriateness of others. Theory is not prescriptive but constructive and descriptive. The whole point of theory is that it addresses areas of knowledge or perception in our lives that we find difficult to talk about (phronesis) (Bergendahl, 1990, p. 186; Gadamer, 1976, pp. xxv and 120) and brings them in contact with the craft of communicating (techne), so that we can try to find articulations adequate to the context of our day to day lives, that will eventually enable us to make decisions about events and to act.

Appropriate action is the broader significance of the world 'proper' as in search for a 'proper meaning' (Hunter, 1989, pp. 169–70). In the humanities, this activity is sharply noticeable in the theoretical debates about gender and race, in groups of people who work within a dominant discourse, a dominant set of common grounds and strategies, that ignores and even represses their different knowledge, have stepped away from opposing the repression to attempting to find ways of articulating the repressed knowledge. The debates about 'class' are particularly difficult to reduce to terms not of opposition but of unarticulated knowledge. Each area needs to establish certain norms of communication for granted in its theory, or literally, it will appear to be talking nonsense. Articulating what is known but not yet articulated is a process involving the speaker, the medium and the audience. If the audience has no *common* ground of communication at all, neither will it be about to understand the significance or sense of the text.<sup>7</sup> At the same time, each area is self-consciously assessing common grounds and attempting new ones, and that reflexivity is not unitary because each one has different guidelines about where to halt and sort out the value of its representations.

Modern science differs from the arts and humanities in that rather than having other human beings or other human-created texts as its referent, it claims the natural world and the constructions that people have made out of it. This difference has resulted in a profound difference in the way that the sciences and humanities communicate to any public audience. For the 'hard' sciences, knowledge is expressed in the natural work (phronesis); it is known from doing the experimentation (techne) and practising scientists are fairly continually reflexive about this engagement with reality by way of assessing the grounds of their scientific method (Gooding, 1989; Rose, 1992). But while the referent participates in this, it does not talk back. Furthermore, communication about this knowledge to a public is second-order articulation about the first-order experiment, and therefore, finds it easy to look for normative and conventional grounds for expression rather than reflexively assessing their appropriateness (Gooding, 1989; Medawar, 1986). The knowledge is supposed

about the world not about a changing communicative society; the communication already at one remove from the real activity.

The attempt at methodology which has come to define modern science in the Western world since the 17th century, can be seen as an attempt to provide precise guidelines that halt the reflexivity that public discussion entails (Shapin, 1994). But this attempt to halt reflexivity is not in aid of evaluating the grounds appropriate to social action. In science, the stress on methodology and on technical experimentation, can be seen as an attempt to question the unsatisfactory area of plausible grounds of scientific argumentation, by taking science *out* of the social demands of communication. The desire of the Royal Society for a pure cognitive language can be recast, not in terms of a concern for the objectivity of language for its own sake so that it can refer to the world precisely and accurately, but for the objectivity of language so that it can represent the experiment exactly. Further, the new science needed a plain language without its own textuality because many Protestant reformers of that early time it wanted, and still often wants, to stand by themselves, as if words were bags to put the reality known in the world experiment into. This denial of textuality is exacerbated from another direction: technology begins to use science to make commercial profit from the 18th to the 19th century. What permits technology to commodify or make money out of science is its ability to represent science mechanically (Henderson, 1986) as a fixed methodology producing 'true' products.

It is a self-defeating agenda: science defines language as inadequate to represent the experience of reality, and suitable only for the reporting of experimental methodology. Hence the public understanding of science is restricted to the clichéd version of scientific prose as a parody of technical rhetorical structure: *propositio, inventio, narratio, conclusio*; and scientific language as rational, analytical and cognitivist (Bazerman, 1987). By closing the door onto society, in an attempt to get rid of the merely plausible, science only succeeded in restricting to physical demonstration its communication about what it was effectively concerned with (Gooding et al., 1989), and in producing a set of representations of its methodology that did not take account of common grounds – which is exactly where the plausible comes from: unassessed habitual agreement.

One of the commentators to have emerged from the debate over the legitimacy of scientific crisis, whose work is pertinent because it appears to applaud and authorise the plausible as community habits of interaction, is Richard Rorty. Rorty rejects textuality or grounds that representation is inadequately referential (1991a, p. 8), and in this duplicates the rejection of rhetorical stance enacted by science when it accepts that written language cannot convey the action of experimentation. He notes that this inadequacy of representation raises insoluble questions about the absolute and the relative, and wishes to replace it with 'habits of acting' and interaction with a 'neighbourhood' (p. 10). It is the case that if a stance is adopted that wants manmade control over the aspects of language that model it as a mirror of the world, then these dichotomies emerge; indeed most of 20th century psychoanalysis engages with this and its other consequences of the stance of fantasy which technically it is (Hunter, 1994). But there are other choices of rhetorical stance such as allegory or dialogism, or a variety of poetic stances that have been built by people who have agreed to look at

appropriateness of the structures of representation rather than dismiss the prob that result from belatedly recognising that language is not exact.

What is curious about Rorty is that instead of choosing another stance, he with this impoverished notion of representation and because of its impoverish chooses to ignore it altogether. His arguments suggest that if representation is drop then metaphysics and epistemology can be replaced by politics, which is to as that metaphysics and epistemology have never had anything to do with pol whereas if rhetorical stance and social context is retained then so is the pol dimension. In effect, in dropping representation, analysis leading to political a is evaded; for Rorty *phronesis* never needs to be articulated – indeed articulati textualising is seen as a reductive and ultimately intolerant act (1991a, p. 25). How as the rejection of rhetoric by science underlines (Hunter, 1984), drop representation, avoiding textuality, simply disguises the fact that all communic involves power relations. Discourses that pretend otherwise, like science, bec ‘the most potent instrument(s)’ of power for persuasion in our society (Ca 1989, p. 161). The reason that Rorty’s position is curious is that the comfor elision into competency that marks exclusive club cultures and their excess, science/technology to liberal/authoritarian regimes (Bauman, 1989, p. 12; R 1991a, p. 44), is something those societies frequently attempt to resist, yet w Rorty offers as a desirable end; and desirable is just what it is. Elsewher characterises ‘habits of acting’ as the private narcissism and public pragmatist (210) that make up the solidarity of club culture. It is however, only the ‘leisure’ ‘civilisation’ of bourgeois liberalism (p. 25) that permits the desires of pr narcissism to remain ‘tolerant’ rather than brutal.<sup>8</sup>

Bourgeois liberalism assumes a large measure of individual autonomy, but it effect, dependent on an institutionalised state structure that creates conditior private autonomy within a highly regulated public practice. Rorty suggests ‘civilised’ culture is autonomous from the institutional regulations of police bureaucrats (p. 26), whereas it can only behave *as if* it were so by refusin become aware of the institutional structure. It is the need for the doublethin simultaneously acknowledging state structure and then forgetting it in order to op as a private pragmatist, that underlines how deeply embedded is this approach in structures it seeks to evade, and which rather surprisingly, aligns it with versio Althusserian state apparatuses,<sup>9</sup> although it owes its primary allegiance to Feyerab The doublethink is here called ‘ethnocentrism’, and is contrasted in an ar specifically on science, with rationalism and its “criteria for success [which laid down in advance” (p. 36); but, in effect, ethnocentrism is simply what hap to rationalism when it forgets the need to look at common grounds, treats the given and does away with textuality. While the reflexivity of scientific experit is its strength, its rejection of the textuality of communication has left it to b increasingly enclosed worlds, small circles of private knowledge exchanged by m of a specialist jargon. These worlds are often valuable but frequently bec alarmingly narrow and reductive and only continue to exist because of instituti support (p. 39).

In contrast, for the arts and humanities, communication in a medium is a order activity. Because they are concerned with human referents which engag the activity of knowing, communication itself becomes the site of knowledge. W



is not to say that there is a necessary connection with society: the further away arts move from public discourse the closer they come to duplicating, in, for example, the 19th century avant-garde, the small circles of private knowledge typical of the sciences (Williams, 1987). But both the arts and the humanities find common ground in the activity of their human texts, in the mimetic performing, repeating, imitating of these texts (Cave, 1986). This is one of the reasons why it is so important for arts and humanities users of computing to understand the constraints of the system of computing by a software text (Goranzon, 1990; Miall, 1990), let alone hardware: they need to know not only what the rules of the software will permit, but how to perform within the textuality within the large critical context (Hunter, 1990). It is in this latter activity that they can reach evaluations, assessments of the knowledge that the interaction between software text and literary text has enabled.

Because the communication of their knowledge is part of that knowledge, the arts and humanities have also traditionally claimed to focus on reflexive social questioning of methodology at the same time as using it, although they have not always carried it out. When they do, they help to maintain the socially immediate criteria by which we know how the common grounds of agreement and representation are being assessed and encourage attempts to be clear about the relations of convention and dominance that are involved and about the relation between articulated and unarticulated knowledge. Such evaluation of stance helps to make clear whether we are seeing new common grounds, rationally extending old ones, breaking up existing common grounds, breaking through existing common grounds, offering different ones, replacing old ones, assessing old ones, using old ones self-evidently or fictionally or artificially and so on.<sup>10</sup>

### **AI and the Representation of Scientific Methodology**

For AI, both articulated and unarticulated knowledge are related to the activity that mimics human intelligence (Goranzon, 1990). Yet because AI conventionally provides a theoretical space for working out the representation of science and technology, most AI currently deals in representation as if it were only cognitive (Wagman, 1991). It is as if it were taking the limited representations of methodology in science and technology for the first-order actions (phronesis) necessary to recognize and imitate (techne) in order to articulate knowledge. Hence it is concerned not with textuality which encourages repetition of an activity that locates (mimesis), finds and assesses the appropriateness of common grounds for all kinds of knowledge perception, but with a stable set of common grounds that evades the implications of their plausible basis and seeks only to represent experimentation or at worst, to the 'success' or 'failure' of that representation. The frustration with this approach is becoming more evident in AI writing itself, with some writers claiming to write without representation at all because it is self-delusory, or to shift it into plural (Brooks, 1991, p. 142; Clancey, 1992, p. 194).

The problems here are specific. First, that if assessment is not self-conscious, old common grounds will be taken as self-evident, yet second, that constant testing is an arbitrary gamesmanship that lets one win or lose but does not offer new common grounds for action. Since the communication of science and the commercial activity of technology (Miles et al., 1988; Molina, 1989) depends upon treating the self-evident

as axiomatic and the unarticulated as irrelevant knowledge, AI is restricted to process of looking for rationally acceptable topoi (Hunter, 1991b) – models, diagrams, analogues, frames (Greene, 1986) – in order to represent. Because of the restriction of cognitivism, Hubert Dreyfus defines a ‘cognitivist’ as a rationalist with a computer (Dreyfus, 1988), AI tends to find these representations in regulative rule-based systems.<sup>11</sup>

This tendency to reproduce the patterns of intensely private theory, and to view within the structures of artificially closed worlds, is what ties AI to fantasy, to the rhetorical stance of technology and industrial capitalism (Johnson, 1989; Hunter, 1991b) which attempts to create isolated worlds that deny larger social and political interaction (Ennals, 1991, p. 41). Yet it is precisely there that assessing the appropriateness of common grounds is not only most difficult but also necessary. The concepts most frequently discussed by people involved in theorising AI are intimately tied to the techniques for building self-defining worlds. The central devices are: algorithmic knowledge (Penrose, 1989), predicative knowledge (Janik, 1991), digital representation, propositional descriptions (Lyotard, 1979), regulative rules (Wittgenstein, 1953; Polanyi, 1967), determinist structures. Each is a part of a doublet that presents its other face, which is frequently offered as the reflexive alternative: mathematical knowledge, tacit knowledge, analogical representation, narrative description, constitutive rules, relativism. What this essay will go on to do is discuss the usefulness of these doublets to AI in an attempt to contextualise its methods and become reflexive about its practices. The discussion here will argue not that the determinate and relativist doublet can be ignored, or that it should be evaded (Rosen, 1991a), but that AI has failed to pursue reflexivity because it was misled into taking the communication of scientific method for the practice, and has as a result, avoided the complexities of representation, particularly of constructing new grounds for unarticulated knowledge and perception. The immediate questions then become: How can representation become self-conscious or reflexive? What is unarticulated knowledge? What can it be for AI? Can it become articulated, and if so, how? How can we halt reflexivity in social appropriateness?

## **AI and the Articulation of Working or Tacit Knowledge**

As the limitations of rationalist, cognitivist approaches have become insistent, there has crept toward the social by introducing contextual factors. The pre-determined common-grounds of the methodologies may satisfy attempts to know fact, information and data, but they fail to ask about knowledge of practice which demands specific context (Winograd and Flores, 1986; Ennals and Gardin, 1990; Ennals, 1991). There is a striking problem here: if practice, or working knowledge of how something is done, is often silent and inarticulable, how can AI represent it? One of the anxieties about working with models provided by AI derives from this awareness that because the models give the illusion of completeness, people working with them will not access the unarticulated practical knowledge that comes from working with the actual world (Cooley, 197X; Goranzon, 1990). But if the representations work by encouraging reflexivity about their limits and exclusions then the user gets access to analogous ways of representing practical knowledge, that are at the same time, ways of participating in that knowledge (Ennals, 1991, p. 71). One result of this latter approach

is the recognition that what people who work with computers need above training (Boden, 1989, p. 166), not just in information representation, but in a set of skills developed in the arts and humanities for representing a much wide of knowledge.

## Dreyfus and Janik

If AI has crept toward context, many theorists of philosophy and epistemology explicitly raised (social) questions about AI and practical knowledge. So w practical or tacit knowledge? The point of agreement among many differen that tacit knowledge is knowledge by the expert, the knowledge of the expertis is learned in practice. In an attempt to illustrate why AI models often beat begi but rarely measure up to experts, Dreyfus outlines a “phenomenology of s behaviour” that moves from novice, to advanced beginner, to competent proficiency, to expertise. In the process he describes a gradual internalisati heuristic procedures so that the proficient performer is able to understand a pro with immediacy and then “must deliberate about how best” to act (Dreyfus, 19 105); however, the expert has internalised both understanding and action, so th [sic] ‘does not solve *problems*. He does what in his experience, has normally w and, naturally, it normally works (p. 106). What is missing from this account i attempt to describe how the ‘intuitive’ expertise differs from cognitive heuri nor does Dreyfus attempt a representation of expertise. The remainder of the illustr he offers ascribes the roots of cognitivism in Western epistemology to Plato, v is ironic, since Plato attempts one of the most enduring representatio non-cognitivist expertise in his accounts of writing, medicine and gardening i *Phaedrus* (Hunter, 1984).

It is precisely this problem with the representation or articulation of exp that is central to the future of AI. Richard Ennals has called it the ‘bottleneck’ development; and Alan Janik, who has been a guiding hand to many comment interested in the implications of practice for AI, makes the possibility of articul the focal point in his definition of the tacit knowledge of expertise. Janik distingu between tacit knowledge that can be articulated but happens not to be, such as secrets or recipes (Janik, 1987, p. 55), and tacit knowledge which cannot be articu even if we wanted to do so. Here he offers the two examples of “certain non-v sensory experience and procedure involved in following a specific rule” (p. 55) elucidation of both kinds of tacit knowledge for labour management and understanding the “limits of the possible” in actual working contexts, is clear helpful. It provides a way of talking about ‘quality’ within the economics of produ (p. 61), as well as the basis for a critique of both Habermas and Gadamer. J describes the irrelevancy of universalist/generalising principles of legitimatio labour-relations working within structures of domination; but he also insists recognition of the limits of any particular community or set of actual condit This is essential if philosophers are to halt the continual reflexiveness typical o cynicism and nostalgia of postmodernism.

However, there is a problem with the example of “impossible to articu knowledge. The case on non-visual sensory experience that is offered is the ‘s of coffee’ (p. 56): because we do not have an articulation adequate to describin

smell so that someone unfamiliar with it could recognise it, we have to use metaphors and analogies. What Janik does not go on to consider is what happens then.<sup>12</sup> In common with agreements about visual representation, there are frequent agreements made by particular groups of people about the adequacy of an articulation representing other sensory experience. People not part of a group will have as a problem recognising visual representation, as they will with representation of other sensory experience. To this extent, the visual is not distinguishable from other senses, although it often appears to be so because technological development and global media have educated a considerable part of the world's population to specific and sophisticated agreements about it. And, to the same extent as the visual, other senses are open to education in representation. Wittgenstein uses the example, the aroma of coffee, to discuss the way that we experience a misdescription (1953, p. 610), but he is far less definite about the impossibility of finding a verbal articulation for it.

More interesting, but also more problematic, is Janik's description of how a constitutive rule-following resists articulation, and this is an argument also informed by Wittgenstein. Janik notes:

Rule following activity entails the kind of knowledge that is only acquired through repetition or practice. It terminates in creative activity as we learn that we have to guess how to continue to follow the example we have been given and, ultimately, as we learn to invent new ways of carrying on...In short it is analogical rather than a digital activity...Moreover, if constitutive rules were known before their application we could never learn to apply rules without more rules...(p. 578).

What is not considered here is the way that constitutive rules tend to become regulative with repetition, with time and with certain social contexts (Conner 1989). The way in which social context effectively brings together both regulative and constitutive, is underlined by Wittgenstein (1953, pt. 179e) when he refers to 'tacit presuppositions' on which any analogical 'language-game' rests. Different from Austin directs the latter part of *How to do things with Words* (1962) at distinguishing between what he calls the constative (descriptive) and the performative (constructive) as part of a larger attempt to distinguish between locutionary (correspondence to the facts), and illocutionary, utterances. He notes that while the constative may be 'right' in the sense of being 'true' to "your knowledge of the facts", the constative and performative may raise questions about 'right' in the sense of whether "this was the proper thing to say" (p. 145). Social context, or "the total speech act situation is the *only actual* phenomenon" (p. 148) which may elucidate the distinction.

Social context is brought together with time and repetition in the consistent coherent history of the process of topical or analogical reasoning documented by the arts and humanities (Hunter, 1991a). A topos or a common ground will be hammered out into language among a number of people attempting to find a place, literally ground upon which to discuss and decide issues. Janik himself uses the phrase "seeing where we are" (p. 57). Once the topos is 'in place' and has provided helpful ground for learning about issues, the sense of its specific appropriateness elides into generalisation. The related procedure of the 'cliche', or 'cliquer' from the clicking of type into a printer's composing stick which allows for multiple repetitions of the same thing, is exactly to repeat an appropriate analogy so often that it becomes regulative, self-evident and banal. At the same time, repetition locates

the foregrounded artificiality of the analogy or 'canonised example' or 'mod 'pattern' (Janik, 1990, p. 50) that is representing practical knowledge, and the an becomes an accepted convention for that knowledge.

The articulation of practice will always be under historical and social pressure; hence the articulations underpinning regulative rules will become inappropriate on new common grounds, genres, canonical models and topoi will encourage and necessitate new articulations. Whether the 'need' is answered is often dependent on financial and economic strictures. Janik hints at the importance of history and its necessity in his criticism of Hubert Dreyfus as a 'phenomenologist' who does not distinguish between constitutive and regulative rule-following. Here Janik says that regulative rules are "only possible on the basis of the former [constitutive rules]" (1987, p. 48), but like Dreyfus he does not expand on the historical process by which the one representation can lead to the other.

## Wittgenstein

Wittgenstein's *Philosophical Investigations* and much of his later writing is precisely concerned with a study of how we can represent "what is not the case" (1953, p. 110), which may account for the considerable number of commentators on practical knowledge who turn to his work. Wittgenstein defines 'philosophy' much like the humanities define 'theory', as discovery and description (pp. 47–48) of what happens when normal, prescriptive rules do not work, when we get 'tangled up in the rules of the language game' (p. 50). Just as teaching and learning need to be done in a way of example and practice (pp. 63–64) rather than by propositions, so must language in providing practical knowledge of reality, not limit itself to conceptual/cognitive representation. The language game is not an explanation but the "primary method" (p. 167). It must provide models as 'comparative' not dogmatic (p. 51); it must understand the need for 'appropriate' words (p. 54). Common grounds are not 'given' simply because they make events more likely (the plausible), but because they are an experiential 'influence on the event', they are part of it, they make it 'more probable' (p. 136).

Underlying Wittgenstein's investigations is a careful critique of the practical cognitive and behavioural psychology contemporaneous with his writing. Frequently the writer distinguishes between human 'sciences' and natural sciences on the grounds that study of the former depends not upon measurement, calculation, instrumental concepts, but upon the external sensory reactions of the subject (p. 151). At the end of the investigations explicitly stating that through experimental methodological approaches that psychology makes to behaviour are a 'conceptual confusion' which results in 'barrenness'. This statement derives from the lengthy section (ii.xi) which discusses the difference between seeing and 'seeing as', and that also provides Janik with a basis for distinguishing constitutive from regulative behaviour in terms of an analogical from digital structure. Evaluation and judgement arise not from explicit rules but from the imitation of a practice by which we learn what to see (Janik, 1987, pp. 52–53). Tilghman (1990) uses this distinction to pursue questions of articulation and to define 'seeing as' as a practical knowing that is "usually an ability to use the appropriate language and/or make appropriate comparisons". Wittgenstein leaves us with an uneasy indication that the socially 'proper'

'appropriate' slide under pressures of time, use and history, into nominalisation. As he notes in the *Blue and Brown Books*, analogies can outstay their appropriate and become misleading (Hunter, 1989)

## Polanyi

The focus on Wittgenstein and away from Michael Polanyi, who is often considered as the writer who opened up a wider discussion of the topic of tacit knowledge, ostensibly because Polanyi is considered 'vague' or even 'mystical'.<sup>13</sup> But this apprehension derives in part from Polanyi's stress on modern science rather than the arts and humanities, which means that he is not immediately concerned with the appropriate representation for communicating to a wide public, but with the appropriate experimental representation which allows the individual to articulate physical reality. In other words he is concerned with a different referent for the unarticulated tacit knowledge, and with a different relationship to the audience. Since AI is a field of knowledge concerned with its own representation, Wittgenstein is highly relevant. However, AI gains much of its impetus directly from scientific discourse, and often takes that discourse as a first-order representation of the knowledge of modern science with the result that it bypasses much of its articulation on regulative rules. Hence Polanyi's careful study of the difficulties of finding a representation for scientific experiment is also highly relevant.

In *The Tacit Dimension* (1967) Polanyi talks about scientific discovery as being guided by the sense of a presence of a hidden reality which discovery 'terminates and satisfies'. In doing so, "it claims to make contact with reality" (p. 24). What is interesting is that Polanyi's sense is congruent with Wittgenstein's on philosophy that discovery is a series of hunches that are recognised as problems in the 'performance' or physical practice of the experiment, by the experimenter/scientist. This sense of the scientist being reflexive by looking for problems rather than solutions, and of science as performance rather than objective inquiry, forms the basis of Lyotard's discussion of science 20 years later, as the central activity of the postmodern condition. But Polanyi carefully notes the long history of this activity. He outlines the problems that this reflexiveness inevitably meets because it lacks connection with the social that a concern with the textuality of representation ensures.

## Doublethink of Postmodern Pluralism: Arbitrary and Absolute or, Private and Systematic

Lack of effective social mediation leaves the scientist grasping after private truth often called beauty,<sup>14</sup> or the enormous satisfaction and pleasure of the system coherence of ideology. The isolated scientist can never know whether the discovery or experimental practice is private truth about actuality, or has in effect, suddenly made sense of a pre-existing system of epistemology. In other words, the scientist can never answer the question of whether the inarticulable tacit knowledge has been (partially) articulated or whether the experiment has just got around to articulating something that no one else had yet bothered with. Any private pursuit of truth ends with this dilemma about the arbitrary or the absolute. Postmodernism engages in the doublethink so well described by George Orwell: of the arbitrary as absolute or the absolute as arbitrary.

There have been attempts to rescue the 'discovery' of science by relocating it firmly away from the cognitivist 'algorithms' of computing, into the 'beauty' and 'mystery' of mathematics. Penrose (1989) attempts this restatement of classic truth and beauty, in order to point out the delusions of the 'strong AI' claim that 'the information content' of individuals "can be translated...intact" (p. 26), that is inadequate "for the description of brains and minds" (p. 402). Penrose understands the view that computing is algorithmic, following rules mindlessly, and incapable of evaluation, judgement of truth, or artistic appraisal; and this is an approach which he has been taken to task (e.g. Sloman, 1992). Here the emphasis is rather on the private and egocentric grounds for the representation involved in his definition of creative mathematics. He claims that the "evaluation of an algorithm cannot be a matter of *conscious awareness*" (p. 447), because the consciousness of mathematical truth is a matter of instantaneous 'seeing' that involves 'inspiration, insight, and origin' and the "strong conviction of the validity of a flash of inspiration...is very closely linked up with its aesthetic qualities" (p. 421). It is suggested that a mathematical truth is not information but a discovery akin to Plato's 'remembering' – that the mathematical concept "is, in a sense *already present* in the mind" (p. 429). It is an absolute.

The reference to Plato and memory is significant because remembering what is 'already present' is with Penrose narcissistically still limited to the isolated individual and does not take on the rather more active sense of re-membering, rebuilding the body, that occurs when the individual comes into contact with society that is offered by Plato. The narcissistic image is the other face of the psychoanalytic vocabulary that recasts beauty as the nostalgic recognition of the self in the symbolic system of ideology where truth is a designed, interconnected network, and is linked up with the cynical recognition of the difference of the self from the symbolic system. Significantly, both narcissism and cynicism depend upon the private individual who is empowered with respect to the institutional structures of the public world. Postmodern plurality is never challenged by itself, only by the different communities that it excludes and which may not be tolerant of its naive arrogance.<sup>16</sup>

Penrose tries to distinguish the individual discovery of the 'already present' truth, from the sense of the brain 'observing itself' whenever a new perception emerges that is offered by systematic coherence. However, the two are the alternative sides of the relativist and the deterministic, the private revolution and the total system, neither of which acknowledge the difficulty of dealing with the real, which cannot be approached without social interaction. Furthermore, once either approach begins to question its own methodology, it cannot cease because there is no self-interest/material/social need that can appropriately interrupt and focus on action (Dreyfus, 1989, p. 82; Rorty, 1991a, p. 39). What often happens instead is that the desire for commodification and capitalism intervenes, turning the scientist or science into fame or technology. Both fame and technology claim to satisfy desire by closing the circle of knowledge, so they immediately turn the constitutive rules of performance into regulative games in which the rules may be completed (Frude, 1989, p. 10). This is the precise way that commodification offers the seductions of power to the speaker and audience. Plato equates it with paying money for love (not sex).

Polanyi attempts to differentiate between this relativist/determinist double and science's apprehended contact with the 'real', by way of the disturbance that the discovery causes, saying that discoveries are 'most real' when they manifest themselves

in the future in the largest range of indeterminate results (1967, p. 24). But Polanyi also realises that the difficulty of distinguishing, in the elation of discovery, between the joy of an appropriate articulation for tacit knowledge and the 'termination satisfaction' of power either individual or systematic, is the condition of a self-validating world of knowledge isolated from society. He adds that the scientist can utter "no...more than a responsible commitment" (p. 78) to the truth, and that the universal claims of science are not deterministic but of "universal *intent*" (p. 7).

In his conclusion Polanyi notes that because the heuristic field of science is stable but problematic, the scientific community must be spontaneously established by self-coordination' (p. 92) or it will lead to a fragmented society "adrift, irresponsible, selfish, apparently chaotic", where each scientist is responsible only to their own small specialisation. These comments neatly describe the effects of postmodernism which Rorty would like to celebrate in the name of bourgeois liberalism; which Gadamer tries to harness into self-coordination while failing to address the dilemma of distinguishing between plausible and probable grounds (between prejudices and traditions), which Lyotard leaves adrift in the melancholy fragments of cynicism and nostalgia: the mirror of Baudrillard's 'simulacrum', and which Habermas would apparently like to address by reintroducing enlightenment ideals.

If tacit knowledge for science is knowledge of the reality of the physical world, for the arts and humanities it is knowledge about the world of human interaction. Any common grounds are shared with other human beings as the basis for all types of social action. Any attempt at new common grounds for unarticulated knowledge must involve other human beings. For the arts and humanities the sense of suddenly sharing a new common ground with other human beings is part of the knowledge being articulated. The articulation and the making public are the same thing, although it could sensibly be argued that the social sciences stress a reality about human beings that is described in the repetition of appropriate representations, and the critical sciences/arts stress the contact with reality that is learned by repeating performed moves in processes of imitation as if they were our own. In both sciences there has been a traditional focus on providing contexts for representations, rather than delineating methodologies, although the 'scientist' drive of the humanities in the 20th century has been to take on methodology enthusiastically. While the approaches have opened up the variety of ways in which we study and understand other human beings, the methodology has frequently been used as an end in itself, falling into the same trap as AI of taking the methodologies themselves as reality to be represented. This has derailed much work from questions of reflexivity into similar consolations of power offered by the commodities of fame and of technology of criticism, with the result that the power has been left without appropriate context. It is this that has led the arts and humanities into ahistorical postmodernism and generated what has come to be known as the 'legitimate crisis' in all disciplines of Western study.

Gadamer, Habermas, Lyotard, and more recently Rorty, offer versions of the same crisis, stemming from the same doublethink of the private and systematic, the arbitrary and the absolute. Gadamer does not deal with the problem of the 'common ground as ideological because he chooses to work from the human animal out (19 pp. 110–111), as if biological necessity will inevitably be more significant than ideological. He provides private reasons for reflexively finding and critiquing common



grounds but not overthrowing them because implicitly he doesn't believe or ever get outside them: there is no discussion of the pressure of contradiction, need to maintain common ground for taking decisive actions, of the problem of competing common grounds, and the problem of not wanting to or not being able to critique the common ground characterised as a 'false consensus'. He distinguishes between 'good' and 'socially successful' argument (1987, p. 194), on the basis of whether the argument is questioned and defended or merely influential. However, this offer of reflexive questioning as the way to distinguishing is not enough: a 'socially successful' argument may be challenged and then defended, while a 'good' argument may have to change with social context. His distinction can in the end only be maintained by calling on 'universals'. Lyotard simply gets rid of the terms necessary and universal, leaving the crisis of legitimation with the incessant reflexive game that shuttles back and forth between the isolated private world and the system or universal.

What is interesting about Lyotard's reductiveness is that it underlines the central issue of evaluation: how to account for, assess, and act on, perceived contact or knowledge of, the real. There is no room for practical or tacit knowledge. Lyotard's version of 'paralogy', plurality of worlds. This is primarily because there is no belief in a contact with the 'real'. All representation operates within paralogical structures of private or systematic worlds. A recent version of this account replicates Lyotard's rhetoric. In a paper analysing the 'hard' representation of science by sociologists of science criticising 'hard' representation of nature, Steven Weber maps out three stages in sociology (1991, p. 44). The first is 'instrumental', the second 'interpretivist' and the third, 'reflexive'. The tripartite categories roughly approximate the divisions in rhetoric between technique, strategy or ethos, and stance. However, what rhetorical study emphasises is that while consciousness of stance encourages reflexivity, it is not automatically something to be proud of, as Weber would have it. Reflexivity does insist on placing the 'object' (knowledge, scientific person) within the analyst's own social forces, but the concept of stance adds that this is not a unitary activity: there are different ends and values involved in social action in the context for an object. The problem is not just 'to become reflexive' but to decide at which point it is appropriate to halt reflexivity in order to take decision-effect social policy.

Here it can be appropriate to refer to the story of Pygmalion. The continually reflexive mind, endlessly deconstructive and sceptical, is implicitly utopian, searching without end for the ideal: that Pygmalion will be able to make Galatea articulate reality perfectly. Yet Pygmalion needed the gods to start and to stop his articulation. To ignore the gods of biology, or any other universal, does not suddenly legitimise Pygmalion's activities by conferring authority on him alone. Indeed, it merely renounces the current rules of thumb about reality, and leaves Pygmalion playing arbitrary games with Galatea within a self-enclosed monomaniac world.

### **3. AI and Representation**

One reason that theorists/philosophers such as Gadamer and Habermas hang on to biological necessities and universals, is that they perceive the need for social ac-

The whole process of articulating tacit knowledge is at the root of the liberatethos of modern science, as it is for the humanities. What goes on in the articulation of tacit knowledge is the making public of the ways in which, or the points at which we have decided to halt reflexivity and take action. In any scientific approach, partly because the communication of experimental work is already at one remove from the actual, and partly because of its methodological emphasis, the moment at which the knowledge derived through practice is articulated it appears to make claims about fixed truth. But what it is in effect doing is making claims on reality, and then intervening in it and changing it. Looking at how knowledge is made public can provide guides to finding the 'appropriate' places to halt reflexivity.

The process of making public any claims on reality does not, for scientific approaches, occur in the medium that the science works. But science still needs public discourse. Hence, the burden of proof and repetition to involve other people in the contact with the real. Articulation of any working knowledge only happens as a result of common practice, otherwise there would be no one to recognise appropriate common grounds. Science specifically, works in small communities where 'common practice' is the first mode of proof (Gooding, 1989). The articulation puts that practice into a representation so that it can then be learned by a wider public, but all too often for reasons previously outlined, the articulation uses strategies that aim at 'truth'. Curiously, this mimics the particular politics of authoritarianism that arose in the 20th century in response to the massive number of newly enfranchised people. The stance is an extension of fantasy that aims to satisfy representation by providing it with the self-validating terms of tautological structures. It frequently fails; it results in force when presented to an audience which does not accept or does not understand, its basic assumptions – as all those who have attempted to assemble something from technical instructions will know. But science need not communicate this way. The knowledge it is about may be learned in quite different ways dependent upon the articulation and particularly upon the stance. AI could be a helpful place to develop and assess more interesting approaches to scientific articulation and the effects it has on scientific practice.

The array of dyads offered by commentators on AI and representation – constitutive/ regulative; mathematical/algorithmic; tacit/predicative and systematic; analogical/ digital; narrative/propositional, etc. – marks each first term as typical of the moment of elation where we think we touch reality. Strictly speaking, the dyads are different in kind. The constitutive or tacit knowledge of working practice is articulated as mathematical, analogical, narrative representations, that may become algorithmic/ digital or propositional, and lead to regulative and predicative/systematic rules. However, there is no necessity for the elision from say analogical to digital, rather there is varying historical and social pressure. For example, if you live in a society with poorly developed practices for reading analogy into the large number of figures of which metaphor, metonymy, simile, allegory and so on, are but figures, then analogy will slip swiftly into digital analogue. Different articulations effectively affect practice in differing ways dependent on this pressure.

Once articulated, practical knowledge need not be merely related to a rule-based system, because the text of articulation can permit the repetition of the performative practice or work that insisted on the articulation in the first place. However, at the same time as it makes it possible, articulated practical knowledge does inhibit

articulation of other practices. Some way of representing will be closed off, others will be opened up, so it becomes necessary to be reflexive about representation, to understand what our position is with respect to it, and how to participate in it, not so that we can engage in the incessant shuttle between the private and the systematic (whether universal, ethnocentric or ideological), but that we can connect the articulated with the unarticulated and take a step to evaluating appropriateness.

To answer the questions of how much and where we are participating in representation, and what any articulation makes it difficult to say, we have to address the common grounds in the practice of that representation. Nearly all the commentators interested in AI and society, whether or not they are concerned with tacit knowledge, discuss these questions of representations. The point in each case is to underline the need for constitutive practice, the first term within each dyad of representational concern, not always followed up, is the need for reflexivity that is the initial step to constitutive practice. But constitutive practice involves an understanding of rhetorical stance which brings together the strategies of reflexivity with its own critique of articulation; it brings together strategy and the material; it focuses precisely on the elision from constitutive to regulative and asks us to be clear about our position, about where to stop reflexivity and take action.

The stance of AI has predominantly been based on the self-enclosed rhetoric of scientific communication. It moves inexorably around the structures of tautology, making that scientific methodology has elaborated, taking common ground as axiomatic, as 'agreement', as 'best fit', as oppositional, as satisfactory, as unnerve and in this way raises peculiarly acutely the problems inherent in fantasy. The stance of fantasy as elaborated above is a way of dealing with unarticulated knowledge in social context, by treating all common grounds as systematic, yet choosing to simultaneously to forget that strategy. Once the system has been defined, it satisfies any desire articulated within its common grounds. It can offer continuous repetition of that satisfaction, it becomes addictive and the compulsive addiction to computer worlds is well-documented.

A recognition of this self-deluding procedure leads R. Brooks to try to completely evade the need for representation in AI. He concludes that:

When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model. (1991, p. 14)

What this reveals is that he takes 'representation' to be explicit and so of course finds it inadequate. Furthermore, using the world's 'own' models necessarily involves an articulation of another representation. As he proceeds, Brooks appears to become aware of this. He becomes insistent upon 'traditional' representation being the first of the problem, and he turns elsewhere inventing in his notion of robotics interest new possibilities for AI. But as Kirsh's commentary indicates, these have their own implicit representations (Kirsh, 1991b, pp. 161-184). Any evasion of representation on the grounds that it is inadequate to the world is tautological because it is based on the concept of a linguistic adequacy which fails by definition. Kirsh points out that the idea of an inadequacy of axiomatic knowledge is also bound up in an essentialism based on egocentricity; the denial of a public and social space (1991a, p. 21). This concern parallels the broader ethnocentricity of Rorty, with his similar evasion

inadequate representation. And Brooks' comments that he is not interested in the significance of the implications of his work, indicates a similar evasion of social concern.

The kind of satisfying repetition enabled by chosen or evaded tautological work is quite different from the repetition enabled by the constitutive practical performance, because the common grounds of performance can never be defined within a fixed system. This realisation forms the centre of W. Clancey's monograph *Model construction operators* (1992), about which he comments:

I realised that many confusions about representation could be resolved if we see them as alternative perspectives on a single 'virtual' formal system...I realised that representation and reasoning procedures that were commonly viewed as different could be related by a shift in visualization. (p. 193).

Yet Clancey's commentary on the way his approaches have developed, almost exactly parallels the shift in scientific concerns with representation from major to the plurality of postmodern design and its attendant problems. Some researchers are turning to alternative ways of agreeing upon grounds for definition and upon grounds for reference (Birnbau, 1991; Kirsh 1991a, b; Smith, 1991). Rather than rushing directly into the plural versions of postmodern representation which, as commentators on representation in science have recognised, are still quite separated from society, perhaps AI could consider using the concepts of rhetorical strategies present in the humanities to pursue alternative ways of agreeing upon the grounds for the representation it articulates.

Were AI able to investigate the constitutive rules of topical and analogical reasoning it might be able to offer science a mode of proof and repetition that reintroduces the phronesis of science into its representation, reintroduces the probable and insists on the textuality of public communication.<sup>17</sup> But if AI is to duplicate in virtual reality the repetition of satisfaction that is the representation of science (by technological means), then it will place itself within the story of Frankenstein. Ulysses, Pygmalion's Galatea, Frankenstein's creation is manmade with naive arrogance; it takes the second-order representation of science and makes it literally real; technology is born and there are no gods to get rid of it. It is helpful to remember that Pygmalion is a myth by man for men, underlining a belief in their own god-like actions, and that Frankenstein and the monster are part of a myth written by a woman looking out of a man's world, and providing a devastating critique of the literalising of fantasy.

## Notes

1. Quote from D Lenat and R Guha, *Building large knowledge-based systems* (Reading, MA, Addison-Wesley, 1989).
2. See, for example, the analysis of the rhetoric of totalitarianism and authoritarianism in the work of nationalism by Gellner (1982), Hunter (1984) and Rowe and Schelling (1991).
3. For arguments in favour of club culture, see Rorty (1991 a,b) and for various critical, if not condemnatory arguments, see Jameson (1990) and Bauman (1989, 1992)
4. Whether or not it is historically accurate to portray pre-Platonic sophistic rhetoric as concerned with an unstable ethos, recent interest in it indicates the anxiety about political structures that have to deal with multicultural, multiracial and multiclass diversity.
5. Much of this work over the past 30 years has been focused on the writing of J Derrida who problematises 'ethos' in "Plato's Pharmacy" (1967), and could be said to have extended the notion of deconstruction specifically to deal with the problems raised by the inadequacy of ethos for an increasingly diverse audience.

6. See the classic formulation of the way technology and experiment are part of reality in P Bachelard *formation de l'esprit scientifique*, Paris, Vrin, 1938.
7. See R Gooding on witnesses and lay observers to early science experiments in Gooding pp. 191–192.
8. See Z Bauman on the ambivalence of tolerance, p. 237.
9. For a helpful version of Althusser's 'Ideology and ideological state apparatuses', see D McD *Theories of discourse* (Oxford, Basil Blackwell, 1986).
10. These are often known as theories of difference, but without social context they have no signifi
11. On not doing this, see Stockinger (1990) and Fargue (1990).
12. This is, of course, the field of poetics; recent discussions of the unarticulated have focused on t language for women's experience (Cixous, 1987) and for the experience of races other than the do (Bhabha, 1983).
13. Janik disagreeing with and quoting Dreyfus and Janik (1990) p. 48.
14. This link may date quite specifically from the 18th century, although the association needs research. G. Cantor (1989) quotes the scientist Robert Millikan saying on an experiment, "E Publish this surely, beautiful", pp. 159–160.
15. J Lacan sets the stage for this movement in the essays delivered during the 1950s and 60s, collec Ecrits (1976).
16. For further commentary on this aspect of the political positioning of the individual, see Bauman t on 'competent consumers', p. 259.
17. Much recent work on the early elaboration of public demonstrations of science indicates that this r of textuality has been present at least since the 17th century; see for example, S Shapin (1988).

## References

- Aleksander, L. (1989). Connectionist systems: Information technology goes brain-like (again!). In Murray and Richardson
- Austin, J. (1962). *How to do things with words*. Clarendon Press. Oxford.
- Bauman, Z. (1989). *Modernity and ambivalence*. Polity.
- Bauman, Z. (1992). *Intimations of postmodernity*. Routledge. London.
- Beetham, D. (1991) *The legitimization of power*. Macmillan. London.
- Benjamin, A., Cantor, G., and Christie, J. (1987). *The figural and the literal: Problems of language and history of science and philosophy, 1630–1800*. Manchester University Press, Manchester.
- Bergendahl, G. (1990) Profession skill and traditions of knowledge. In Göranson and Florin.
- Bernstein, R. J. (ed.) (1985). *Habermas and modernity*. Polity.
- Bhabha, H. (198X). Interrogating identity. *Identity Documents*, 6.
- Birnbaum, L. (1991). Rigor mortis: a response to Nilsson's 'logic and artificial intelligence. *Artificial Intelligence*, 47.
- Boden, M. (1989). Artificial Intelligence: opportunities and dangers. In Murray and Richardson.
- Boden, M. (1989). The meeting of man and machine. In Murray and Richardson
- Boden, M. (1990) Artificial intelligence and images of man. In Ennals and Gardin.
- Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, 47.
- Cantor, G. (1989) The rhetoric of experiment. In Gooding et al.
- Cave, T. (1988). *Recognition: a study in poetics*. Clarendon Press. Oxford.
- Charolles, M. (1990) Logician analysis and discourse analysis. In Ennals and Gardin.
- Christensen, K. (1986). Technological decisions are moral decisions. In Geiss and Viswanathan.
- Cixous, H. (1988) 'Coming to writing' and other essays. D Jenson (ed). translation S Cornell. Harvard University Press. Cambridge, MA.
- Clancey, W. (1992). Model reconstruction operators. *Artificial Intelligence*, 56.
- Clancey, W. (1993). Notes on 'heuristic classification'. *Artificial Intelligence*. 59.
- Connerton, P. (1989). *How societies remember*. Cambridge University Press. Cambridge.
- Cooley, M. (197X). *Architect or bee? The Human/technology relationship*. S. Cooley (ed.). Hand and B Publications. Slough.
- Cooley, M. (1987). Creativity, skill and human-centred systems. In Göranson and Josefson.
- Dear, P. (ed.) (1991). *The literary structure of scientific argument*. University of Pennsylvania Press. Philadelphia.
- Doran, J. (1989). Distributed artificial intelligence and the modelling of socio-cultural systems. In Murray and Richardson.
- Dreyfus, H. (1988). The Socratic and Platonic basis of cognitivism. *AI & Society*, 2. 107.
- Ennals, R. (1990). Interpretation and code-breaking. In Ennals and Gardin.

- Ennals, R. (1991). *Artificial intelligence and human institutions*. Springer-Verlag. London.
- Ennals, R. and Gardin J.C. (eds.) (1990) Interpretation in the humanities: perspectives from artificial intelligence. *Library and Information Research Report 71*. British Library.
- Fargue, J. (1990). Remarks on the interrelations between artificial intelligence, mathematical logic and humanities. In Ennals and Gardin.
- Feyerabend, J. (1975). *Against Method*. Verso, 1988.
- Foucault, M. (1976). *Language, counter-memory practice*. Cornell University Press. Ithaca.
- Frude, N. (1989). Intelligent systems off the shelf: the high street consumer and artificial intelligence. Murray and Richardson.
- Gadamer, H-G. (1976). *Reason in the age of science*. Translated by FG Lawrence, MIT Press, 1981, London.
- Gardin, J-C. (1990). Interpretation in the humanities: some thoughts on the third way. In Ennals and Gardin.
- Geiss, G. and Viswanathan, N. (1986) *The human edge: information technology and helping people*. Hawthorn Press. New York.
- Gellner, E. (1982). Nationalism and the two forms of cohesion in complex societies. In Proc British Academy LXVIII. Oxford University Press. Oxford.
- Gill, K.S. (1987). Artificial intelligence and social action: education and training. In Göransson and Josefson.
- Golisinski, J. (1990). The theory of practice and the practice of theory: sociological approaches in the history of science. *Isis*, 81.
- Golisinski, J. (1992). *Science and public culture*. Cambridge University Press. Cambridge.
- Gooding, D., Pinch, T. and Schaffer, S. (eds.) (1989). *The uses of experiment: studies in the natural sciences*. Cambridge University Press. Cambridge.
- Göransson, B and Florin, M. (eds.) (1987). *Knowledge, skill and artificial intelligence*. Springer-Verlag. London.
- Göransson, B and Florin, M. (eds.) (1990). *Artificial intelligence, culture and language: on education and work*. Spinger-Verlag. London.
- Greene, J. (1987). *Memory, thinking and language*. Methuen. London.
- Habermas, J. (1971). *Theory and practice*. Translation J Viertel. Beacon Press, 1973. Boston.
- Habermas, J. (1985). Reply to Rorty. In Bernstein.
- Habermas, J. (1987). *The philosophical discourse of modernity*. Polity.
- Hunter, L. (1984). *Rhetorical stance in modern literature*. MacMillan. London.
- Hunter, L. (1989). *Modern allegory and fantasy*. MacMillan. London.
- Hunter, L. (1990). Fact-information-data-knowledge: databases as a way of organizing knowledge. *Journal of Literary and Linguistic Computing*, 5. 1.
- Hunter, L. (1991a). McLuhan's From choice to archetype. In L Hunter, 1991.
- Hunter, L. (1991b). Remember Frankenstein: rhetoric and artificial intelligence. *Rhetorica*, XI, 4.
- Hunter, L. (ed.) (1991). *Toward a definition for topos: essays on analogical reasoning*. MacMillan. London.
- Hunter, L. (1992). Social contexts for methodology. Paper presented to the International Society for Linguistic Computing. Oxford University, (forthcoming printed version).
- Jacob, P. (1990). What is interpretation?: a philosophical view. In Ennals and Gardin.
- Jameson, F. (1984). Foreword. In Lyotard, 1979.
- Janik, A. (1987). Tacit knowledge, working life and scientific 'method'. In Göransson and Josefson.
- Johannesson, K.S. (1990). Rule following and intrasitive understanding. In Göransson and Florin.
- Johnson-Laird, P. (1989). Human experts and expert systems. In Murray and Richardson.
- Kirsh, D. (1991a). Foundations of AI: the big issues. *Artificial Intelligence*, 47.
- Kirsh, D. (1991b). Today the earwig, tomorrow man? *Artificial Intelligence*, 47.
- Lacan, J. (1979). *Ecrits*. Translated by A Sheridan. Tavistock. London.
- Lyotard, J. (1979). *The postmodern condition: a report on knowledge*. Translated by G Bennington and B Massumi. Manchester University Press. Manchester (1986).
- McDonnell, D. (1986). *Theories of discourse*. Basil Blackwell. Oxford.
- Medawar, P. (1964). Is the scientific paper a fraud? *Saturday Review*, August 1st.
- Miall, D. (1990). An expert system approach to the interpretation of literary structure. In Ennals and Gardin.
- Miles, I, Rush, J, Turner, K, and Bessant, J. (1988). *Information horizons: the long-term social implications of new information technologies*. Edward Elgar. Aldershot.
- Molina, A. (1989). *The social basis of the microelectronics revolution*. Edinburgh University Press. Edinburgh.
- Murray, I. and Richardson, J. (eds.) (1989). *Intelligent systems in a human context: development, implications and applications*. Oxford University Press. London.
- Nordenstam, T. (1990). Language and action. In Göransson and Florin.
- Ostberg, G. (1990). How to make materials data systems useful for designers. In Göransson and Florin.
- Penrose, R. (1989). *The Emperor's new mind: concerning computers, minds and the laws of physics*. Clarendon University Press. Oxford.
- Polanyi, M. (1967). *The tacit dimension*. Routledge and Kegan Paul. London.

- Pravitz, D. (19XX). Tacit knowledge – an impediment for AI? In Göranson and Florin.
- Rorty, R. (1985). Habermas and Lyotard on postmodernity. In Bernstein.
- Rorty, R. (1989). *Contingency, irony, and solidarity*. Cambridge University Press, Cambridge.
- Rorty, R. (1991a). *Objectivity, relativism, and truth*. Cambridge University Press, Cambridge.
- Rorty, R. (1991b). *Essays on Heidegger and others*. Cambridge University Press, Cambridge.
- Rose, H. and Rose, S. (eds.) (1976). *The political economy of science*. MacMillan, London.
- Rose, S. (1973). Can science be neutral? *Proc Royal Institute*, 45.
- Rotblat, J. (ed.) (1982). *Scientists, the arms race and disarmament*. Taylor and Francis, London.
- Rouse, J. (1987). *Knowledge and power: toward a political philosophy of science*. Cornell University London.
- Rowe, W. and Schelling, V. (1991). *Memory and modernity*. Verso, London.
- Searle, J. R. (1990). Cognitive science and the computer metaphor. In Göranson and Florin.
- Shapin, S. (1988). The house of experiment. *Isis*, 79.
- Sloman, A. (1992). The Emperor's new mind. *Artificial Intelligence*, 56.
- Smith, B. (1991). The owl and the electric encyclopedia. *Artificial Intelligence*, 47.
- Stockinger, P. (1990). Logicist analysis and conceptual inferences. In Ennals and Gardin.
- Tilghman, B. R. (1990). Seeing and seeing as. In Göranson and Florin.
- Wagman, M. (1991). *Artificial Intelligence and human cognition: A theoretical intercomparison of realms of intellect*. Preager, London.
- Wilcock, J. (1990). A critique of expert systems, and their past and present use in archaeology. In Ennals and Gardin.
- Williams, R. (1987). Paper presented at the Linguistics and Literature Conference, Strathclyde Unive
- Winograd, T. and Flores, F. (1986). *Understanding computers and cognition. A new foundation for de* Addison Wesley, Norwood, NJ.
- Wittgenstein, L. (1953). *Philosophical investigations*. Translated by G Anscombe. Basil Blackwell, O: (1967).
- Woolgar, S. (1989). Why not a sociology of machines? An evaluation of prospects for an association bet sociology and artificial intelligence. In Murray and Richardson.
- Woolgar, S. (1991). The turn to technology in social studies of science. *Science, Technology and Hu Values*, 16 (1).
- Young, R. M. (1989) Human interface aspects of expert systems. In Murray and Richardson.
- Ziman, J. (1982). Basic Principles. In Rotblat.

---

*Correspondence and offprint requests to:* Lynette A.C. Hunter, Senior Lecturer, University of Le School of English, Leeds LS2 9JT, UK.

# The Contribution of Tacit Knowledge to Innovation

Jacqueline Senker

*Science Policy Research Unit, University of Sussex, Brighton, UK*

**Abstract:** Tacit knowledge is widely acknowledged to be an important component of innovation but such recognition is rarely accompanied by more detailed explanations about the nature of tacit knowledge, why such knowledge is significant, how it becomes codified or whether it may be limited to codification. This paper attempts to fill some of the gaps, drawing on a study of university/industry links in three emerging technologies. It concludes that tacit knowledge, which can only be transmitted through personal interaction, will continue to play an important role in innovation. This derives from a variety of reasons, but most significant are the complexity of systems and the emergence of new technologies.

**Key words:** Tacit knowledge; Innovation; Automation; Biotechnology; Advanced ceramic; Parallel computing

## Introduction

Studies of innovation, technology transfer and technology diffusion identify tacit knowledge as an important component of innovation (e.g. Dosi, 1988; Roser 1976, 1982). None of these discussions provide a satisfactory definition of tacit knowledge or give a detailed, systematic account of its role in technological innovation. Nor do they provide guidance on how its importance may differ according to the industrial sector or technology being studied or how firms acquire it.

This paper is an attempt to begin to fill some of these gaps. It will discuss the methods developed by firms to capture tacit knowledge of a scientific and technological nature which is generated both within the company and that arising in external sources. It will also identify the company activities where tacit knowledge makes an important contribution to innovation and consider the processes by which tacit knowledge becomes codified.

Much of the material arises from a recent study of university/industry links in three emerging technologies – biotechnology, advanced engineering ceramic parallel processing – which investigated all the knowledge flows used by companies in the course of innovation (Senker & Faulkner, 1992). Results from this study will be presented after proposing a working definition for tacit knowledge and reviewing