

Empirical Modelling and the Foundations of Artificial Intelligence

Meurig Beynon

Department of Computer Science
University of Warwick, Coventry CV4 7AL, UK

Abstract. This paper proposes Empirical Modelling (EM) as a possible foundation for AI research outside the logicist framework. EM offers principles for constructing physical models, typically computer-based, by construing experience in terms of three fundamental concepts: observables, dependency and agency. EM is discussed in the context of critiques of logicism drawn from a variety of sources, with particular reference to the five foundational issues raised by Kirsh in his paper *Foundations of AI: the Big Issues* (AI, 47:3-30, 1991), William James's *Essays on Radical Empiricism* (Bison Books, 1996), and the controversy surrounding formal definitions for primitive concepts such as metaphor and agent that are recognised as fundamental for AI. EM principles are motivated and illustrated with reference to a historic railway accident that occurred at the Clayton Tunnel in 1861.

The principal thesis of the paper is that logicist and non-logicist approaches to AI presume radically different ontologies. Specifically, EM points to a fundamental framework for AI in which experimentally guided construction of physical artefacts is the primary mode of knowledge representation. In this context, propositional knowledge is associated with phenomena that are perceived as circumscribed and reliable from an objective 'third-person' perspective. The essential need to incorporate subjective 'first-person' elements in an account of AI, and the role that commitment plays in attaching an objective meaning to phenomena, are seen to preclude a hybrid approach to AI in the conventional sense.

1 Introduction

More than ten years have elapsed since McDermott's celebrated renunciation of logicism in AI first appeared [59]. The status of neat and scruffy approaches to AI remains controversial, and there has been limited progress towards the two complementary goals that might make the most decisive impact on the argument:

Goal L ("The Logicist Goal"): Develop sophisticated symbolic models with powerful applications.

Goal NL ("The Non-Logicist Goal"): Identify general principles for application development outside the logicist framework.

By way of illustration, Goal L would be served if the aspirations of Lenat and Feigenbaum's experiment in knowledge representation [57] were to be realised, and Goal NL by the discovery of general principles for constructing reactive systems of agents sufficient to account (in particular) for the achievements of Rodney Brooks and his collaborators at MIT [32, 31].

A cynic might argue that neither of these goals has significant academic or commercial implications. Whether or not logicism delivers significant practical applications, the logicist view of AI is firmly entrenched in the curriculum of computer science, and will remain so whilst there is perceived to be no academically respectable alternative to mathematical foundations based on logic and rules. And whether or not there are any fundamental principles to account for the achievements of scruffy AI, those techniques that are most effective in practice will attract commercial interest and exploitation.

More considered reflection suggests that progress towards one or other of the goals set out above potentially has serious implications in both academic and commercial terms. As Brooks argues in [31], AI has been very influential in shaping the development of computational paradigms and knowledge representation techniques, and its agenda is increasingly relevant to modern computing applications. The limitations of traditional foundations for computer science are becoming topical in many key areas. Recent contributions on this theme include, for instance, West's discussion of hermeneutic computing [76], Wegner's proposals for extensions to the Turing model of computation [75], and the analysis of information systems development by Hirschheim *al.* [52]. Related concerns include the future for database concepts beyond relational and object-oriented frameworks, and Fred Brooks's reiteration of his call for principles with conceptual integrity to address the problems of software development [30]. A logicist framework that fails to engage with the agenda of modern practical computing calls into question the integrity of AI and computer science as academic disciplines. Computing practice that has no coherent guiding principles is unlikely to deliver well-engineered products or to exploit the full potential of new technology.

The aim of this paper is to consider the potential of Empirical Modelling (EM), developed by the author and his collaborators at Warwick over several years, as a broader foundation for AI and Computer Science. (See our website: <http://www.dcs.warwick.ac.uk/pub/research/modelling> for further details of the Empirical Modelling Project.) By way of clarification, it should be noted that the term 'agent' has a distinctive meaning in EM that has been developed largely independently of the associations that 'agent-oriented' now has in Computer Science. By way of disclaimer, this paper aims to expose a fundamental difference in preoccupation between the logicist and non-logicist perspectives, and should be interpreted as questioning the potential significance rather than the intrinsic merit and interest of logicist researches. For instance, automatic techniques for truth maintenance and belief revision are a legitimate way to ameliorate the effects of adopting a closed-world perspective, but this does not address the fundamental problem raised in section 2.3 below.

The paper is in three main sections. Section 2 contrasts logicist and non-logicist perspectives on intelligence with reference to a typical IQ puzzle and to the analysis of a historic railway accident. Section 3 introduces EM principles and techniques, and illustrates their potential significance for railway accident investigation. Section 4 discusses the new foundational perspective on AI that EM affords with particular reference to the work of William James on Radical Empiricism, of David Gooding on the empirical roots of science, of Mark Turner on the roots of language and of Rodney Brooks on robotics.

2 Perspectives on Intelligence

In [55], Kirsh discusses the foundations of AI with reference to five issues:

- Core AI is the study of conceptualization and should begin with knowledge-level theories.
- Cognition can be studied as a disembodied process without solving the grounding problem.
- Cognition can be effectively described in propositional terms.
- Cognition can be studied separately from learning.
- A uniform architecture underlies virtually all cognition.

Kirsh identifies these as assumptions typically associated with a logicist viewpoint. EM promotes an alternative viewpoint on intelligence. In particular, it takes a different stance on each of these five foundational issues.

The essential distinction concerns the way in which a system is construed to operate. As discussed in Kirsh [55], the logicist aims at a mathematical structure of objects, functions and relations close enough to the real world for a system to achieve its purposes, and construes the system as “acting as if it were inferring”. In EM, in contrast, a construal makes more explicit reference to human agency, can have a more subjective character, and be more loosely concerned with specific goals. A central idea of EM is that physical artefacts are needed to communicate such construals, but its general import can be expressed as: “so far as I/we can judge from previous experience, and subject to exceptional behaviour for which there is no pre-conceived explanation, the system is acting as if it were composed of the following family of autonomous agents, each responding to the following observables, and exercising the following privileges to change their values in the context of the following dependencies between observables”.

A construal in EM has a number of key features:

- It is empirically established: it is informed by past experience and is subject to modification in the light of future experience.
- It is experientially mediated: the interaction in which each agent engages is represented metaphorically via a physical artefact, typically computer-based.
- The choice of agents is pragmatic: what is deemed to be an agent may be shaped by the context for our investigation of the system.

- It only accounts for changes of state in the system to a limited degree: the future states of the system are not circumscribed, there may be singular states in which conflicting values are attributed to observables, and there are no guarantees of reliable response or progress.

Construals in logicism and in EM are associated with radically different ontologies and epistemologies. This ontological distinction is highlighted when, as is sometimes appropriate, EM is used to develop models of a logicist character. The critical point in this development is the point at which the developer shifts perspective from “my experience so far suggests that this pattern of interaction between agents occurs reliably and this appears to conform to the experience of others also” to “a particular behavioural pattern within the system is described objectively and precisely by the following logical model whose construction is predicated upon the assumption that this pattern occurs reliably”.

Two examples of interpretations of intelligence will be used to highlight this difference in perspective, and to motivate the more detailed discussion and analysis that follows.

2.1 A Classic Intelligence Test

The problem posed in Box 1 illustrates one popular view of intelligence that has much in common with the logicist perspective as portrayed in [55]. It is drawn from a publication by Mensa, a society whose membership comprises people with a high “intelligence quotient”.

The Captain of the darts team needs 72 to win. Before throwing a dart, he remarks that (coincidentally) 72 is the product of the ages of his three daughters. After throwing one dart, he remarks that (coincidentally) the score for the dart he has just thrown is the sum of the ages of his daughters. Fred, his opponent, observes at this point that he does not know the ages of the Captain’s daughters. “I’ll give you a clue”, says the Captain. My eldest daughter is called Vanessa. “I see”, says Fred. “Now I know their ages.”

Box 1. A Mensa Intelligence Test

The solution to this problem centres on the fact that factorisations of 72 into 3 factors are disambiguated by the sum of factors but for the pair of factorisations:

$$72 = 3 * 3 * 8 = 6 * 6 * 2.$$

By observing that he does not know the ages of the daughters, Fred discloses to the solver that one or other of these factorisations of 72 is the required one. (Note that, to make his observation, Fred does not need to know—as we as solvers do—that no other pair of factorisations of 72 into three yields the same

sum, since he knows that the Captain has scored 14.) When he knows there is an *eldest* daughter, he knows that the ages of the daughters are 3, 3 and 8.

This puzzle illustrates several ingredients of logicism discussed in [55]. The problem is contrived around a mathematical model in the poser's mind. The casual and artificial way in which the abstract problem is placed in a real-world context echoes the modularity of 'inventing conceptualizations' and 'grounding concepts' presumed in logicism [55]. Embodiment plays a contrived role in the problem. The issue of psychological realism is not addressed. It is assumed that Fred exercises instantaneous—or at least very rapid—inference skills on-line, whilst “knowing the ages of the daughters” is an abstract concept, unconnected with being able to associate an age with a daughter who might turn up at the darts match. Nor is indexicality respected. In order to draw any inferences, a single Mensa-like persona must be imposed on the agents in the puzzle (the Captain and Fred) and on the poser and solver also.

The remarkable thing about problems of this nature is that the IQ-literate reader adopts the conventions of the problem poser so readily. Why should we regard problem-solving of this nature as intelligent? Perhaps because it involves being able to see through the contrived presentation to make ingenious abstract inferences, discounting the commonsense obstacles to deduction (*cf.* Naur's analysis of logical deduction in Sherlock Holmes stories [63]: “truth and logical inference in human affairs is a matter of the way in which these affairs are described”).

To some degree, facility in making abstractions is a quality of intelligence. Some commonsense facts about the world must be taken for granted to make sense of the problem. For example, a game of darts takes place on such a timescale that the ages of the children are fixed for its duration. 14 is a legitimate score for one dart. Yet the puzzle is posed so artificially that it is almost a parody of intelligence.

A complementary mental skill is far less well-represented in logicism. This is the ability to transpose the problem imaginatively so as to disclose the implicit presumptions about the relationship between the abstract and the real-world elements. Imagination of this kind can subvert the intelligence test. A suspension of disbelief is needed in supposing that the Captain and Fred are mathematically adept and sober enough to factorise 72 in their heads whilst simultaneously taking turns at darts, or that Fred determines the ages of the children because of an inference rather than because he remembers Vanessa's age. In some contexts, especially where creativity or design are concerned, such questioning of the premises of a problem is essential, but it is out-of-place in the world of Mensa problems. The intended world model is closed and preconceived.

The Mensa problem is an example of the kind of challenge that might be addressed by an intelligence inference engine. It might not be easy to meet, as it involves some meta-level reasoning. This is illustrated by the fact that if Fred said he knew the ages of the daughters before he was told the name of the eldest, no inference could be drawn.

Though logicism is not primarily concerned with artificial tests of intelligence of this nature, it can be seen as construing intelligence in similar terms. It involves

establishing a formal relationship between the world and a logical model similar to that between the mathematical model and the darts match scenario, such that intelligent behaviour can be viewed as if it were inference of the kind used in solving the intelligence test.

Empirical Modelling techniques address the broader view of intelligence that encompasses creativity and imagination. They are not particularly well-suited for exercises in inference masquerading as commonsense problems, but have direct relevance to real-life scenarios in which abstract explanations are sought.

2.2 The Clayton Tunnel Railway Accident

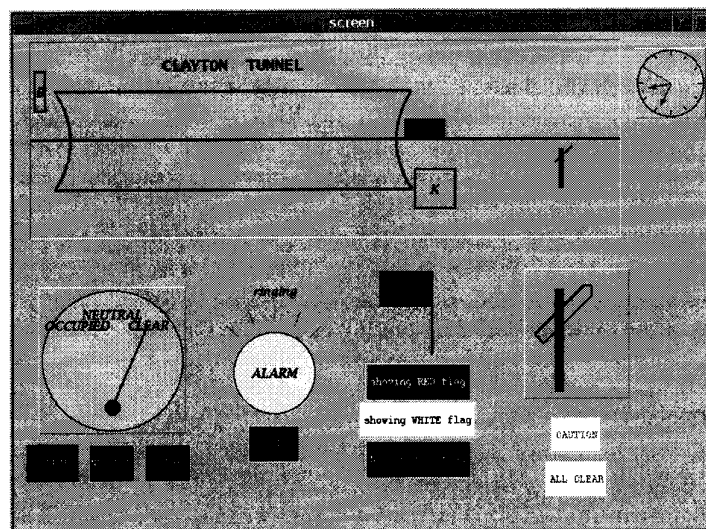


Fig. 1. Signalman Killick's view of the Clayton Tunnel

The following discussion refers to a 19th century railway accident [67] that is described in Box 2 and illustrated in Figure 1. In analysing the accident (e.g. as in conducting an accident inquiry), the significance of embodiment is particularly clear. To assess the behaviour of the human agents, it is essential to take account of psychological and experiential matters. How big was the red flag? How was it displayed? Did the drivers and signalman have normal sight? How far away could oncoming trains be seen? These are perceptual matters, which taken in conjunction with knowledge about how fast trains travelled and how closely they followed each other, help us to gauge the performance of human agents. There are also conceptual matters, to be considered in the light of the training given to drivers and signalmen. It is reasonable to expect that a responsible driver can

The Clayton Tunnel Disaster

August 25th 1861

Three heavy trains leave Brighton for London Victoria on a fine Sunday morning. They are all scheduled to pass through the Clayton Tunnel—the first railway tunnel to be protected by a telegraph protocol designed to prevent two trains being in the tunnel at once. Elsewhere, safe operation is to be guaranteed by a time interval system, whereby consecutive trains run at least 5 minutes apart. On this occasion, the time intervals between the three trains on their departure from Brighton are 3 and 4 minutes.

There is a signal box at each end of the tunnel. The North Box is operated by Brown and the South by Killick. K has been working for 24 hours continuously. In his cabin, he has a clock, an alarm bell, a single needle telegraph and a handwheel with which to operate a signal 350 yards down the line. He also has red (stop) and white (go) flags for use in emergency. The telegraph has a dial with three indications: NEUTRAL, OCCUPIED and CLEAR.

When K sends a train into the tunnel, he sends an OCCUPIED signal to B. Before he sends another train, he sends an IS LINE CLEAR? request to B, to which B can respond CLEAR when the next train has emerged from the North end of the tunnel. The dial at one end of the telegraph only displays OCCUPIED or CLEAR when the appropriate key is being pressed at the other—it otherwise displays NEUTRAL.

The distant signal is to be interpreted by a train driver either as *all clear* or as *proceed with caution*. The signal is designed to return to *proceed with caution* as a train passes it, but if this automatic mechanism fails, it rings the alarm in K's cabin.

The accident

When train 1 passed K and entered the tunnel the automatic signal failed to work. The alarm rang in K's cabin. K first sent an OCCUPIED message to B, but then found that train 2 had passed the defective signal before he managed to reset it. K picked up the red flag and displayed it to Scott, the driver of train 2, just as his engine was entering the tunnel. He again sent an OCCUPIED signal to B.

K did not know whether train 1 was still in the tunnel. Nor did he know whether S had seen his red flag. He sent an IS LINE CLEAR? signal to B. At that moment, B saw train 1 emerge from the tunnel, and responded CLEAR. Train 3 was now proceeding with caution towards the tunnel, and K signalled all clear to the driver with his white flag.

But S had seen the red flag. He stopped in the tunnel and cautiously reversed his train to find out what was wrong from K.

Train 3 ran into the rear of Train 2 after travelling 250 yards into the tunnel, propelling Train 2 forwards for 50 yards. The chimney of the engine of Train 3 hit the roof of the tunnel 24 feet above. In all 23 passengers were killed and 176 were seriously injured.

Box 2. An Account of the Clayton Tunnel Railway Accident

interpret a red flag as a signal for danger, and make this inference at the speed of thought (*cf.* the implausibly rapid inferences that Fred must make in his darts match). The process of identifying and actively checking the state of the signal also has a conceptual component.

Issues of this nature have to be viewed with reference to the particular environment, such as the weather conditions. In this context, whether the speed of the train was “too fast” is a matter of pragmatics rather than mathematics. The need to think in egocentric indexical terms is self-evident. None of the human agents has a comprehensive view of the system. Without at least being able to acquire some representative experience of what signalman Killick’s task involved, it is hard to make a fair judgement about his degree of responsibility for the accident, and to assess the relevance of his having worked for 24 hours at a stretch.

In the railway accident scenario, unlike the Mensa problem, the interaction between conceptual worlds and the real world is very subtle. Ironically, the practical measures designed to protect against the dangers of a breakdown in the tunnel also generated the conceptual framework that led to the disaster. Driver Scott’s decision to reverse his train arose from the fiction that a train may have broken down in the tunnel ahead. Had he had another misconception, such as that Killick had waved a white flag, there would have been no accident, and normal operation would shortly have been resumed. In the real world, there are degrees of physical interaction between trains that fall short of the catastrophe that actually occurred, some of which might even have entailed no disruption to the railway system. It is hard to envisage how logicist models could address the range of ways in which what is informally viewed as inconsistency can be manifest. Drastic collocation of trains is a particularly striking example of embodied inconsistency. After this event, there is, in some sense, *no longer a model*.

2.3 A Logicist Model of the Railway Accident?

Kirsh [55] suggests that a theory of AI is concerned with specifying the knowledge that underpins a particular cognitive skill. On this basis, accounting for the Clayton Tunnel Disaster is an exercise of significant intrinsic importance that can be seen as a challenge for a theory of AI. This exercise involves understanding the contributions made by all the human agents in the accident scenario. As part of this process, it would be necessary to place the accident in a generic context, so as to see the actual events in relation to normal operation, and to explore the probable outcomes had the circumstances been different. For instance, there are closely related scenarios in which no accident occurs, or the crash is less violent, and there are myriad factors that could have had a significant influence, such as the reaction times of drivers and signalmen, the effectiveness of braking on the trains, and the geography of the tunnel.

If our objective is to understand the Clayton Tunnel Railway accident in these terms, there appear to be significant problems in constructing a logicist model. To observe logicist principles, it seems that the goal of understanding the accident should lead to the identification of a closed-world model that encompasses the accident scenario and is adequate for all purposes of explanation. Modern railway practice demonstrates that—at least in principle — a closed-world model can be effective in this role, accounting for the highly complex interactions in the railway system within a robust generic conceptual framework.

There are three challenges in particular that are met in conceiving railway system operation in closed-world terms. They are concerned with obtaining guarantees, so far as this is possible, on the following points:

- All human activities are framed around objective knowledge and skills.
- All significant operations are based on highly reliable assumptions.
- Practice does not depend on the specific features of particular environments.

In the analysis of the Clayton Tunnel accident, it is hard to see how to construct a logicist model to meet these requirements.

The need to deal with first person concerns. One possible subgoal for an investigator might be reconstructing the mechanics of the accident. A mathematical model could be developed in terms of such factors as the mass, position, velocity, acceleration, braking efficiency of the trains and friction and gradient in the environment. In this model, agency would manifest itself as changes in acceleration due to manipulation of the throttle and brake.

An alternative model might be aimed at reconstructing the sequence of significant events. This could be built around an analysis of the protocols for interaction between the signalmen and the drivers, e.g. using a mathematical model for concurrency such as process algebra or calculus. Such a model would register the communications between the human agents as abstract events, and enable their possible patterns of synchronisation to be analysed.

From each perspective, the result is a self-contained closed-world model of the accident. That is to say, both models can be developed to the point where, relative to their subgoal, there is apparently no need to make further reference to the physical context in which the accident took place. In accounting for the crash, the mechanical model can give insight into the influence of technological factors and perhaps supply objective information about the train drivers' actions. The protocol model can likewise clarify what communication took place, and help to assess its significance.

In practice, both perspectives are too deficient in psychological terms to be helpful to an accident inquiry in making judgements about responsibility. Both models create objective "third person" accounts that help to clarify exactly what an external observer might have seen, and put this observation in the context of other possible scenarios. Neither gives us insight into how the experiences of the human agents and the physical embodiments of mechanical agents contributed to the accident.

To construct a logicist model that is adequate for understanding the railway accident would certainly require more sophisticated mathematics. What form should such a model take? It would have to model agents so as to take sufficient account of mechanics and how communication between agents is synchronised. It would also have to characterise the interactions between agents in propositional terms in a way that took sufficient account of psychological factors.

The need to deal with provisional and unreliable insight. Understanding the railway accident involves construing the disaster in the context of day-to-day operation of the railway system. This process of construal has no counterpart in the context of the IQ test above. The Mensa problem is posed with a particular construal in mind: we have to assume that the Captain and Fred act on the basis of inference, oblivious to other commonsense factors, such as personal knowledge Fred might have of the Captain's family. In contrast, the construal of the accident admits no 'right answer'. In this respect, it is far more representative of the challenge to knowledge representation involved in intelligent system design.

In considering how the accident is construed, it is important to recognise the scientific and cultural prejudices that can operate. It may appear obvious that we know what agencies and observables need to be taken into account. It is presumably irrelevant whether Killick had a moustache, or wore a red shirt, or that it was Sunday rather than Saturday morning. From an empiricist standpoint, such concerns are not to be absolutely dismissed. Had Killick worn a red shirt, it might have prevented the driver from spotting the red flag in time to give an acknowledgement. There were doubtless many people who speculated on whether the disaster was an act of divine retribution on those who sought pleasure or conducted work on the Sabbath.

It may be argued that the operation of the railway system was predicated on a particular construal, but this in itself is no justification for adopting the same construal in analysing the accident. No doubt the conduct of people during the Great Plague of London was guided to some degree by a construal of how disease spread and could be avoided. Contemporary medical knowledge leads us to analyse the events retrospectively from an entirely different perspective.

It may be tempting to suppose that modern science can be made adequate to the task of construing the entire context of the accident in closed-world terms. Extreme forms of logicism seem to combine a narrow reductionism with a blind faith in the power of propositions to frame the world. It is clear that ergonomic issues to do with human interaction played a part in the Clayton Tunnel Disaster, but our insight into such issues is even now far from being a science fit for the logicist. Nor does the success of modern railways require any such insight: it relies on confining railway system operation to territories of knowledge that are empirically safe.

The need to deal with the particular context. In considering the accident scenario, it is often necessary to speculate on the precise characteristics of the environment for the accident. Sufficient detail has been retained in the account of the accident given above to convey the impression of the richness of the context surrounding the crash. The trains apparently leave Brighton just a few minutes apart; Killick is fatigued; the trains are heavy; it is a Sunday morning. These details may or may not be relevant. Whatever details we include, it seems that language cannot do justice to what we need to know when we probe the circumstances of the accident.

Did Killick have to leave the cabin in order to wave the flag? What was the exact distance between the signal and the cabin, and how much longer would it have to have been for Scott to see the flag? Was Scott supposed to acknowledge seeing the flag? Did his train have a whistle? All these issues require reference to the real situation, and are concerned with the specific characteristics of the particular time and place.

The explanation of particular events can also invoke observables in ways that cannot be preconceived. In the particular scenario of the Clayton Tunnel crash, the signalman needed to know whether the driver—several hundred yards away in the tunnel—had seen the red flag. Perhaps other accident scenarios in which there was no violation of agreed practice would throw up different examples of rogue observables that were never considered by the designers of the protocols or the pioneers of railway and communications technology.

From the above discussion, modelling in the logicist tradition is seen to be intimately connected with identifying contexts in the world that are stable with respect to preconceived patterns of interaction. Validating that such a context has been identified is a pragmatic and empirical matter about which no absolute guarantees can be given. The observables that feature in these worlds, though not necessarily statically predetermined, have to come and go according to preconceived patterns. The agents that operate in these worlds must perform their actions in a manner that respects preconceived integrity constraints. These are the characterisations of *closed worlds* and *circumscribed agency*.

3 Empirical Modelling

The preceding discussion argues the need for an alternative to logicism as a framework for modelling. Accident investigation demands something other than closed-world modelling. In particular, it suggests a specific agenda: modelling from a first-person perspective, with partial and provisional knowledge, and with reference to a specific context. To respect the need to consult the world in the process of model-building, the modelling process should also be situated: it should take place in or *as if* in the context of the situation to which it refers. Empirical Modelling, here introduced and illustrated with reference to the Clayton Tunnel Accident scenario, has been conceived with this agenda in mind.

3.1 Orientation

The context for the Empirical Modelling Project is supplied by what Brödner [28] has identified as a conflict between two engineering cultures:

One position, . . . the “closed world” paradigm, suggests that all real-world phenomena, the properties and relations of its objects, can ultimately, and at least in principle, be transformed by human cognition into objectified, explicitly stated, propositional knowledge.

The counterposition, . . . the “open development” paradigm . . . contests the completeness of this knowledge. In contrast, it assumes the primary

existence of practical experience, a body of tacit knowledge grown with a person's acting in the world. This can be transformed into explicit theoretical knowledge under specific circumstances and to a principally limited extent only . . . Human interaction with the environment, thus, unfolds a dialectic of form and process through which practical experience is partly formalized and objectified as language, tools or machines (i.e. form) the use of which, in turn, produces new experience (i.e. process) as basis for further objectification.

This conflict has both abstract and practical aspects and significance. Brödner attributes "huge productivity problems and failures of AI attempts" to the dominant influence of the closed world paradigm, and adds that "what appears to be a philosophical struggle turns out to be of the highest practical relevance". The conflict is not confined to the "neat vs. scruffy" debate in AI. It is also manifest in Computer Science as a tension between principles and pragmatism that is a source of several unresolved controversies: declarative vs. procedural programming; relational vs. object-oriented databases; formal vs. informal methods of software development.

Three key problems, drawn from different areas of computing, have had a seminal influence on our research:

Is there a universal framework for multi-paradigm programming?

Birtwistle *al.* (1967) [26] introduced the object abstraction to represent programming as a form of modelling. Backus (1979) [6] argued the need for a history sensitive mode of programming with the virtues of declarative programming. Neither programme has generated an entirely satisfactory programming style and their objectives do not seem to be convergent. The modern agenda for computing has to address paradigms for more general applications such as parallel programming, end-user programming and visual programming. This concern is reflected in trends towards varieties of agent-oriented programming [77, 61] and the use of spreadsheets to aid interaction and interpretation in environments for end-user programming [62].

What principles are needed to address complex systems engineering?

Brooks [29] expresses scepticism about most of the current techniques to support the development of large software systems and contends that we have yet to understand the essence of the problem. Formal methods (such as Chandy and Misra [36]) are effective for closely circumscribed problems. Cohen and Stewart [38] identify fundamental limitations that are encountered in rigorous mathematical modelling for complex systems. Fashionable pragmatic approaches to software development in an object-oriented idiom (such as Rumbaugh [68]) at some point have to make an uneasy transition between objects as real-world representations and as programming abstractions. Harel's response to Brooks's challenge [50, 49] invokes the computer both as machine (in commending formal operational semantics) and as instrument (in advocating the use of visual formalisms).

What paradigm for data modelling can support modern applications?

Kent [54] devotes an entire book to the problems that beset the classical database models, and tentatively concludes that there is probably no adequate formal modelling system for representing information on computers. Codd's relational model [37] offers a formal approach that has had a profound impact on commercial database systems. New requirements (knowledge-based systems for design and Integrated Project Support Environments [33], databases for graphics and multi-media, interfaces via direct manipulation, spreadsheets or constraint techniques [40]) have exposed the limitations of the pure relational model. The conflict of cultures pervades the current controversy [5, 39, 72] concerning the relative merits of relational and object-oriented database models. This controversy highlights the need for alternative methods of modelling that associate form and content in new ways.

In each of these problem areas, there is controversy surrounding formal and pragmatic approaches. Our thesis is that fundamental progress in solving these problems can be made only by resolving Brödner's conflict of cultures, developing fundamental principles to complement the closed-world culture. This motivates a radical change in perspective on computer-based modelling.

The Empirical Modelling Project combines abstract investigations and scholarship with practical development of software tools and case studies. EM is a proposal for modelling in an open development paradigm that has emerged from our extensive investigation of principles and case-studies directed at solving the three key problems. The choice of the epithet *empirical* is suggested by the fact that features of a model are typically determined incrementally in the manner of experimental science, and that circumscribed closed-world models can only be derived through explicit acts of commitment on the part of the modeller. Over the last ten years, well over a hundred students have had experience of EM, of whom many have contributed to the research through project work at both undergraduate and postgraduate level. The scope of EM is indicated by the diversity of the notations and software tools we have developed, by the wide range of case studies in modelling that have been addressed and by the many areas of application represented. It is this empirical evidence that informs the discussions which follow.

3.2 Empirical Modelling Principles

The main principles and tools of EM will be discussed and sketchily illustrated with reference to the Clayton Tunnel railway accident. This model is a case-study currently under development by Pi-Hwa Sun, a research student in the Empirical Modelling research group. Details of the tools and notations used to construct the model are omitted, and the emphasis is on the conceptual processes surrounding its construction. For more technical details, the interested reader may consult the EM website and other references cited in [4, 9].

EM is concerned with representing the processes that lead to the discovery of concepts. It differs from a logicist approach in its emphasis upon how concepts are discovered in a psychological sense (*cf.* [55]). In EM, the discovery

process relies upon embodiment in an essential way, and artefacts are seen as indispensable for its representation. The experiential intuitions that inform the construction of such artefacts are here described informally. Practical experience is perhaps the best way to gain a deeper appreciation of EM principles.

The important intuitions on which EM draws are the experience of momentary state (as in “the current situation”), and that of an identifiable pattern of state transitions (as in “a phenomenon”). In the context of the Clayton Tunnel illustration, Figure 1 depicts a particular situation. A phenomenon might be “a train passing through the tunnel”; another might be “a train approaching the tunnel whilst the alarm is ringing”. In EM, an artefact is used to model experimental interaction in a situation, with a view to identifying and construing phenomena associated with this situation.

Construal in EM is relative to the egocentric perspective of a particular agent. Whereas most computational modelling is aimed at realising a system behaviour, the primary focus of EM is on modelling the way that an agent’s construal of a situation develops and how subsequently the conception of a system may emerge. The computer model serves to represent a situation, and transformations associated with the contemplation of this situation. In this context, the computer is being used not to compute a result but to represent a state metaphorically, in much the same way that a physical artefact (such as a scale model, or VR reconstruction of a historic building) can be used as a prototype. The term ‘computer artefact’ is used to convey this emphasis.

The interpretation of computer artefact adopted here is unusual, and merits amplification. It derives from inviting the human interpreter to view the computer as a physical object open to interaction, observation and experiment in abstractly the same way as any other physical object in our environment. Such a view contrasts with the conception of a computer as negotiating input and output to a preconceived schema for interpretation, and in order to perform a preconceived function. This contrast is much sharper than is suggested simply by considering what are often termed the non-functional aspects of the computer operation, such as speed, user convenience and visual effect. The computer artefact is experienced without reference to specific function, and its state is not to be conceived as meaningful only in relation to a predefined abstract pattern of behaviour (e.g. as in the states of a finite state machine). The meaning and significance of the state of the artefact is instead to be acquired through a primitive process of conflating experiences of the artefact and of the external world (*cf.* the *blending* to which Turner refers [73,74]). In this negotiation of meaning, there is no necessary presumption that transitions between states in the artefact reflect familiar objective external behaviours. Rather, like a physical object, the artefact manifests itself in its current state, and my conception of this state is informed by my previous experience, expectations and construal of the situation. By this token, changes to the state of the computer artefact reflect what the human observer deems to be the case: for instance, that one-and-the-same object is now in a different state, or that I now take a different view of this one-and-the-same object.

The framework for construal in EM can be illustrated with reference to an investigation into the Clayton Tunnel Accident. In interpreting the operation of the railway system in the vicinity of the Clayton Tunnel, the investigator will need to identify many different agents and construe the system from their perspectives. The most important perspective is that of an external observer. Initially, the focus is upon how interaction with the model is shaped so as to imitate the experience that an investigator can or in principle might be able to get from conducting experiments in the context of the actual railway system. It will subsequently be clear that the same principles that guide the investigator in developing a construal can also be applied to its constituent agents.

The idea of 'contemplating a particular situation' is illustrated in Figure 1. It is appropriate to think of the investigator as engaged in situated modelling, so that the computer artefact depicted in Figure 1 is placed in the actual environment it is meant to represent. The modelling activity is intended to ensure that the current state of the computer artefact is a good metaphor for the current situation as conceived by the designer. The criterion for goodness is part of the EM concept. What matters is that there is a perceived similarity between the artefact and its referent, a similarity that is recognised through observation and interaction with both (*cf.* Dreyfus's view [43] that human cognition works by 'similarity recognition').

Realism and real-time modelling are not the significant issues. The principal concern is whether observation of the system can be successfully construed: is there in the abstract a way to account for any observed and conceivable changes in state in the system? By way of illustration, Newtonian mechanics can be an excellent way to construe the motion of a train, but it does not of itself deliver a photorealistic real-time animation. In EM, the role of the artefact is to model the way in which the investigator's construal of the system evolves. Changes in the state of the system and changes in the investigator's view of the system are equally significant. Changes of both types are represented in the artefact, and are only distinguished through interpretation. For example, a change to the state of the artefact depicted in Figure 1 could reflect a change in the situation (e.g. the movement of a train), or a change in the investigator's understanding (e.g. the realisation that the location of the signal was inaccurately recorded, that the resetting of the signal was influenced by the weight of the train, or that the colour of Killick's shirt needed to be taken into account).

Fundamental abstractions for EM from a psychological perspective.

EM offers certain principles that guide the analysis of the real-world situation and the construction of its metaphorical representation. The psychological plausibility of these principles is important. Many issues raised by Kirsh in [55] are significant here, and these will be highlighted in the exposition. For instance, in a convincing account of intelligence, the identification of an object—in the generic or particular sense—cannot be taken for granted. There should be a difference between regarding an agent as having 'a symbol in a declarative' and assuming

it to have a concept. A psychologically convincing account of knowledge must offer principles for determining how we extend our concepts to new domains.

EM does not attempt to address the explicit mechanisms by which conceptual developments of this nature are shaped in a human agent's mind. For instance, no consideration is given to the kind of learning processes that are described by neural networks, and might conceivably be persuasive models of brain function. EM simply acknowledges the fact that objects come to be recognised, that concepts are developed and that connections between different situations are established as a result of repeated observation, interaction and experiment. The aim is to develop computer artefacts that can represent the implications of these processes faithfully.

The most elusive but fundamental aspect of the EM approach is its emphasis on modelling a state or situation. This is not to be interpreted as referring to abstract computational state, but to something resembling a 'state of mind' that derives its meaning from a relationship between a human agent and an external focus of interest and attention. This emphasis accords with Kirsh's concern about the embodiment of cognitive skills: "The real problem must be defined relative to the world-for-the-agent. The world-for-the-agent changes despite the world-in-itself remaining constant.". Two of the most important aspects of capturing states of mind are:

- respecting the exceptionally rich and often unexpected associations between different situations in transitions between states of mind
- faithfully registering what is directly apprehended as opposed to what might in principle be accessible.

By way of illustration, an accident investigator might well conceive all kinds of variants of the situation in Figure 1 from within one state of mind. Suppose that the brakes on Train 3 had failed, that Killick had mislaid the red flag, that Train 1 had broken down in the tunnel, that Train 2 had whistled as it reversed from the tunnel, that a different protocol or different railway tunnel had been involved. The significant feature of these variants is the looseness of their relationship to each other: they are not necessarily part of one and the same passage of observation of the railway system (e.g. "the brakes did not fail"); they may not be possible behaviours of the actual system (e.g. "Train 2 was not equipped to whistle"); they can involve transposing events into a totally different context.

The importance of correctly associating observations within one state is also illustrated in the accident scenario. The synchronisation between displaying and seeing the red flag matters crucially. 'Seeing the red flag' and 'recognising the potential hazard' ahead are indivisibly linked in Scott's experience. The analysis of the accident would be influenced if this communication of danger were embodied in a different way (*cf.* "stop if the number displayed on the signalman's card is prime").

There are teasing philosophical issues to be addressed in this connection. What is objective and subjective about the synchronisation of agent actions? In [69], Russell poses a conundrum that concerns establishing the time at which

a murder on a train was committed from valid but inconsistent testimony about synchronisation of events by observers on and off the train. What is the distinction between percept vs. concept? The psychological subtlety of this issue is well-illustrated by this extract from Railway Regulations of 1840 [67]: “A Signal Ball will be seen at the entrance to Reading Station when the Line is right for the Train to go in. If the Ball is not visible the Train must not pass it.”. Such an injunction to respond to what is *not* perceived only makes sense in the context of an expectation that the ball might be seen.

An appropriate philosophical perspective for EM will be considered later. In practice, EM takes a pragmatic stance. Where a logicist model has to address the matter of inconsistency and incompleteness of knowledge explicitly, if only by invoking meta-level mechanisms, EM aims at faithful metaphorical representation of situations as they are—or are construed to be—experienced. There is no expectation that EM should generate abstract accounts of phenomena that are complete and self-contained. In resolving singularities that arise in interpreting its artefacts, there is always the possibility of recourse to the mind that is construing a phenomenon, and to further experimental investigation of the phenomenon itself.

Basic concepts of EM: construing phenomena. The basic concepts of EM are *observable*, *dependency* and *agency*. In the first instance, it is essential to interpret these concepts as egocentrically defined: they are the elements of a particular agent’s construal of its experience, and are appropriately described with reference to personal commonsense experience.

An observable is a characteristic of my environment to which I can attribute an identity. An observation of an observable returns a current value. ‘Current value’ here refers to the value that I would “as of now”—that is to say, in my current state of mind—attribute to the observable. An observable may not always be present, but may disappear, and perhaps later return.

The state of the world for me, as of now, is represented by a collection of observables with particular values. Observables can be physical or abstract in nature: the corner of the table, the volume of the room, the status of my bank account, my ownership of a house.

I might be able to see an observable, or sense it directly in some other fashion, I might have to perform an experimental procedure to determine its current value, or consult an instrument, I might need to invoke social or legal conventions, I might need to use an acquired skill.

Observables are organised in my experience because they are present and absent at the same time, as potential agents, and because their values are correlated in change, through patterns of dependency. Dependency patterns are fundamental to the perception and recognition of observables, and determine when they can be deemed to have integrity as an object. Dependency relations need not respect object boundaries.

Observables, dependency and agency are the focus for two activities: an analysis of my experience, and the construction of a computer artefact to represent this experience metaphorically.

In analysing my experience, I adopt a stance similar to that of an experimental scientist. Repeated observation of a phenomenon leads to me to ascribe identity to particular characteristic elements. To some extent, this attribution stems from the perceived continuity of my observation (e.g. this is the same keyboard that I have been using all the while I have been typing this sentence), but it may stem from a more subtle presumption of conjunction (e.g. this is the same keyboard I was using last week, though I have not been present to confirm this), or another conceptual continuity (as e.g. when I have bought a new computer: that was and this is my keyboard). The integrities that can be identified in this way are observables.

Because the characterisation of observables in EM is experiential and empirical, it is open to a much wider interpretation than a conventional use of the term. When driver Scott sees the red flag, there is no physical perception of the danger of entering the tunnel—indeed, there is no immediate physical danger to be perceived. Nonetheless, the context for displaying the red flag has been established indirectly with reference to expected experience. Danger, of itself invisible—even absent, is present as a conceptual observable concomitant with the red flag. To construe the accident, the investigator must take account of the fictional obstruction in the tunnel that Scott infers when the red flag is seen. And, to deconstruct Scott's concept yet more comprehensively, though Scott could not see even a real obstruction in the tunnel, yet an extrapolation from his recollected experience potentially traces the path from the mouth of the tunnel to the point of collision with this invisible imaginary obstacle.

The idea of dependency is illustrated in the concomitance of 'red flag' and 'danger' as observables. Other examples of dependencies include: the electrical linkage between the telegraphs, whereby the state of a button in one signal box is indivisibly coupled to the state of a dial in another, the mechanical linkage that enables Killick to reset the distant signal, and the mechanism that causes the alarm to sound whilst the signal has not yet been reset.

Dependencies play a very significant part in the construal of a phenomenon. They are particularly intimately connected with the role that invoking agents plays in accounting for system behaviour. A dependency is not merely a constraint upon the relationship between observables but an observation concerning how the act of changing one particular observable is perceived to change other observables predictably and indivisibly. This concept relies essentially upon some element of agency such as the investigator invokes in conducting experiments—if perhaps only "thought experiments"—with the railway system. In empirical terms, dependency is a means of associating changes to observations in the system into causal clusters: the needle moved *because*—rather than simply at the same time as—the button was pressed.

In investigating a phenomenon, dependency at a higher level of abstraction associates clusters of observables into agents that are empirically identified as

instigators of state-change. In a commonsense interpretation of the railway scenario, the agency is taken so much for granted that it may seem perverse to probe its psychological origins, but there are good reasons to do so. The observables that I—in the role of external observer—introduce in construing a phenomenon may enable me to describe the corporate effect of many interacting agents, but there is a proper distinction to be made between my observables and theirs. There will also be typically be certain actions that cannot be attributed to any identifiable agent (e.g. “acts of God”, such as a landslide in the tunnel). My status as observer is reflected in the passive mode in which my construal is expressed in terms of observed actions, possibly attributed to agents, and their perceived effect upon the system state.

My construals are potentially personal, subjective and provisional. What I understand to be the current state will change subject to what kind of phenomenon I am investigating. Judgements about observables, dependency, agency and integrity are pragmatic and empirical matters, about which I can presume no absolute knowledge. By way of illustration, the status of observables associated with trains that have been involved in an accident is obscure. For some purposes (spreadsheet update, timeless experiment), I may be uninterested in how long it takes for the current value to be registered or to be determined. Dependencies amongst observables in the current state reflect the character of my interaction with the environment: e.g. ‘nothing’ can intervene in the updating of a spreadsheet; in a certain context buying a house and signing a document are indivisible by convention; a vehicle will knock me over whether or not I can experimentally determine its exact speed ‘in time’.

The above characterisation of construals is the central abstract contribution of EM, to be discussed in a broader philosophical context below. The central practical contribution of EM concerns the construction of artefacts that are intimately connected with developing construals. This is the focus of the next subsection.

Basic concepts of EM: constructing artefacts. In EM, construals cannot be adequately represented using a formal language: they must be represented by physical artefacts. This representation relies on the perceived correspondence between states and interactions with the artefact, as mediated by its own observables, and those associated with the situation to which it refers.

In practice, the process of construing phenomena is closely bound up with constructing artefacts of just this kind. Construal and artefact construction are symbiotic processes that are interleaved and may even be conflated. Devices that are used to demonstrate the integrity of an observable (e.g. an electrical current) evolve into devices that can associate a value with an observable (e.g. an ammeter), and then become themselves an integral part of a system (e.g. a dynamometer in an aircraft). As these historical examples illustrate, not all artefacts used in the process of construal have been computer-based, but their construction has been restricted by the recalcitrance of physical objects. The viability and topicality of EM stems from the fact that modern computer-related

technology can be the basis for artefacts whose characteristics are no longer so tightly constrained.

Construal in EM can be viewed as associating a pattern of observables, dependencies and agents with a given physical phenomenon. EM techniques and tools also serve a dual role: constructing physical artefacts to realise given patterns of observables, dependency and agency. A key role in this construction process is played by dependency-maintenance that combines the updating mechanism underlying a spreadsheet with perceptualisation. One technique for this involves the use of definitive (definition-based) notations [24].

A definitive notation is used to formulate a family of definitions of variables (a definitive script) whose semantics is loosely similar to the acyclic network of dependencies behind the cells of a spreadsheet. The values of variables on the left-hand side in a definitive script are updated whenever the value of a variable that appears on the right-hand side is updated. This updating process is conceptually atomic in nature: it is used to model dependencies between the observables represented by the variables in the script. A visualisation is typically attached to each variable in a script, and the visual representation is also updated indivisibly when the value of the variable changes. Definitive notations are distinguished by the kind of visual elements and operators that can be used in definitions.

Definitive scripts are a basis for representing construals. In typical use, the variables in a script represent observables, and the definitions dependencies. A script can then represent a particular state, and actions performed in this state can be represented by redefining one or more variables in the script or by introducing a new definition.

The use of two definitive notations in combination is illustrated in Figure 1. One notation is used to define the screen layout and textual annotations, the other to maintain simple line drawings. By using such notations, it is easy to represent the kinds of dependencies that have been identified above. For instance, the dial displays *occupied* whilst the appropriate button is depressed.

If a phenomenon admits an effective construal in the sense introduced above, we can expect counterparts of the transitions that are conceived in exploring the system to be realisable by possible redefinitions in the artefact. In practice, the possible redefinitions do not respect semantic boundaries. For instance, in Figure 1, they may relate to modifying the visualisation (e.g. using a dotted line to represent the track in the tunnel), emulating the actions of agents in the scenario (e.g. resetting the signal), or fantasising about possible scenarios (e.g. changing the location of the signal). This accords with the view of the investigator as resembling an experimental scientist, who, within one and the same environment, can select the phenomenon to be studied, decide upon the viewpoint and procedures for observation, adjust the apparatus and develop instruments.

In practical use, a definitive script can be used judiciously so that all interaction is initiated and interpreted with discretion by the human investigator. For the script to serve a richer purpose than that considered in Naur's account of *constructed models* [63], there must be interaction and interpretation that is

not preconceived. Ways of framing particular modes of interaction that are less open-ended are nonetheless useful. For example, the actions that are attributed to agents need to be identified, and the different categories of action available to the investigator discriminated.

A special-purpose notation, named LSD, has been introduced for this purpose. (The LSD notation was initially motivated by a study of the Specification and Description Language SDL—widely used in the telecommunications industry—hence its name.) The manner in which an agent is construed to act is declared by classifying the observables through which its actions are mediated. This classification reflects the ways in which real-world observables can be accessed by an experimenter. Certain observables can be directly observed (these are termed *oracles*), some can be changed (*handles*), but this change is subject to observed dependencies (*derivates*) and is generally possible or meaningful provided that certain conditions hold (such conditional actions are expressed as a *protocol* that comprises *privileges* to act). It may also be appropriate for a construal to take account of attributes associated with the experimenter (*states*). For instance, the status of certain observations and actions may be affected by the experimenter's location.

An LSD account of an agent can be used in a wide variety of contexts. It can represent what I personally can observe and change in a given situation. Alternatively, it can express what I believe to be the role of an agent other than myself, either from my perspective or from its own. In an appropriate context, it can be also used to specify an agent's behaviour (*cf.* the LSD Engine developed by Adzhiev and Rikhlinisky [3]). These three perspectives on agency are discussed in more detail in section 4.

When construing a complex phenomenon, the presence of several agents leads to potential ambiguity about which perspective is being invoked. For this reason, LSD accounts do not necessarily lead directly to operational models of phenomena. It is not in general possible to develop a faithful computer model of behaviour that can be executed fully automatically; the intervention of the modeller in the role of super-agent is needed to emulate non-deterministic interaction, to resolve ambiguity about the current state of the system, and to arbitrate where the actions of agents conflict. Special tools have been developed for this purpose: they include the Abstract Definitive Machine (ADM) [21], and the distributed variant of the Eden interpreter [23] that has been used to generate Figure 1.

3.3 Characteristics of the EM Construal Process

There are many important respects in which the principles of EM, as described above, engages with the fundamental issues raised by Kirsh in [55]: it is first-person centred, it is not primarily language-based, but experientially-based; it involves embodied interaction and experiment; it addresses conceptualization in psychological terms; it is concerned with intentionality and meaning rather than logical consequence.

As Kirsh remarks: logicists see “inventing conceptualizations” and “grounding concepts” as modular. The most fundamental shift in perspective in EM concerns the nature of the relationship between the artefact and the phenomenon it represents. To say that the artefact metaphorically represents the phenomenon suggests an abstract conceptual correspondence in the spirit of Forbus, Gentner [44, 46], Campbell and Wolstencroft [35]. What is entailed is quite different in character: a correlation between two experiences, one of which is gained through experiment in the world, and the other through experimental redefinition in the script. The presence of dependencies between observables is the psychological mechanism by means of which this correlation leads to a perceived correspondence between observables in the artefact and in the world.

Notice that this process can only be appreciated from a first-person perspective. Only I have simultaneous access to experience of the artefact and the world. This accords with the account of metaphor that is given by Turner in [74]. In its most primitive terms, metaphor is a blending of two experiences within one mind, not the abstract process of establishing a correspondence between abstract structure that is analysed in depth in [35].

In combination, dependency and interaction expose identities and drive the conceptualization process. The way in which pictorial elements in Figure 1 are linked through dependencies in the script is crucial in being able to connect them with the world. What is more, the psychological process of making these connections depends upon being able to exercise a powerful form of agency: that of being able to perform or invoke actions similar to those that an experimenter might perform to test hypotheses about the identity and status of observables (*cf.* Smith’s remark [70] that “agents are what matter for semantical connection”). It is to be expected that pressing the button affects the dial; that applying the brake will slow the train down; that setting the signal to caution will cause the driver to brake. Such experiments lead us to introduce new concepts and observables: for example, to recognise the need to consider the precise point at which the engine triggers the treadle to reset the signal, and to introduce a visual counterpart to the artefact.

A proper appreciation of the experiential basis of the correspondence between computer artefact and real-world referent underlies the distinction between a definitive script and a family of predicates. It is only by virtue of this relationship that it makes sense to regard the variables in a script as referring directly to particular external observables, and the script itself as representing a specific state. In this way, EM addresses Kirsh’s concern [55]: that model theory is a theory of logical consequence, not of intentionality or meaning, and that it does not single out one model or family of models as the intended models. In keeping with the proposals of Smith [70], the relationship between form and content in EM is altogether more dynamic and more intimate than in model theory. Sculpting the form-content relation that binds the artefact to its referent is an integral part of the EM process.

In conjunction with its LSD interface, a definitive script is well-suited to modelling the environment in terms of actions that are afforded, and agents’

dispositions to behave. In this context, the use of dependency-maintenance in mediating the effects of agent actions to other agents and to the external observer is very significant—it circumvents the “ugly semantics” that, as Kirsh observes, stem from the fact that “in stating the role an informational state plays in a system’s dispositions to behave we characteristically need to mention myriad other states” [55].

It is also possible to see how EM can address “tuning the perceptual system to action-relevant changes” both from the perspective of the investigator, and from that of other agents, though practical development of EM is still required to realise its full potential in this respect. This is a strong motivation for developing tools to support the design of artefacts that provide an interface that make better use of human perceptual and manipulative skills. Issues of this nature arise in designing definitive notations. For instance, in devising a definitive notation for geometric modelling, only an empirical evaluation can determine what sort of functional dependencies in a script are best suited to giving the user control of geometric features. Exploring the potential of computer-based technology in constructing instruments in this way is another area of application for EM techniques (*cf.* [47]).

A framework that deals with such experiential matters effectively must still respect the important caveats about displacing logical concepts from an account of intelligence that are also raised in Kirsh [55]. For example: how can the physical fact that the pen is on the desk be seen as the structured facts |the pen|^|is on|^|the desk|? how can ‘the pen is on the desk’ and ‘the pen is matte black’ be seen to entail ‘the matte black pen is on the desk’?

Within the EM framework, it is possible to record conceptual relationships that, though they cannot be explicitly perceived, express the expected results of experimental procedures. Naively, the object is a pen because you can write with it; it is on the desk because it can be lifted from the desk but moves with the desk in a characteristic way. The informality of such criteria do not perturb the empiricist, for whom such assertions enjoy no absolute status of truth in a closed world. The pen may be replaced by a pen-shaped sculpture that has been welded to the table when my back is turned. Interpreting a red flag as a sign of danger illustrates how such conceptual mechanisms can be used.

As for conjoining observations made in the same state, this is the most commonplace activity in constructing a definitive script. To illustrate that yet more sophisticated logic necessarily has sometimes to be invoked in scripts, it is only necessary to consider the complexity of the interlocking mechanisms that were constructed in the railways of the early twentieth century. To represent the indivisible relationships between levers, signals and track points in such a system would involve intricate logical dependencies. Similar considerations apply to the general mathematical functions that can arise on the right-hand side of a definition. The process of surveying land was radically transformed by the discovery and tabulation of trigonometric functions that could be used to reduce the number of explicit measurements needed to make a map.

What is not in keeping with EM principles is the indiscriminate use of logical and functional abstractions that is illustrated in the scenario of the Mensa problem, and in impure variants of logic and functional programming: the agent that is presumed to know the logical implications of the elementary propositions it knows, no matter whether this is computationally feasible; negation as failure; lazy evaluation as a general-purpose semantics for interaction. The account of heapsort in [10] demonstrates how procedural issues can be appropriately addressed using EM principles, and also illustrates appropriate use of relatively sophisticated logical concepts (such as whether the heap condition is valid at a node) within a definitive script.

It remains to consider two foundational issues raised by Kirsh: the extent to which “the kinematics of cognition are language-like” and that “learning can be added later”.

The distinctive qualities of EM as a non-logicist approach stem from the representational significance of the computer artefact. Changes to the artefact can record the adjunction of new observables or the identification of new dependencies. Whilst the interaction with the artefact is open to extension and revision in this manner, the semantics of the model is fluid and arguably cannot be expressed in propositional terms. Modification of the artefact offers a way of representing learning “as a change in capacities behaviourally or functionally classified” [55]. This enfranchises elements of the learning process that are non-linguistic in character, and that are arguably concerned with private processes of incremental understanding that can only be represented with the help of an artefact (*cf.* [9]).

To illustrate the kind of learning activity that can in principle be supported using EM, consider two educational simulations that could be readily derived from the artefact depicted in Figure 1.

One such simulation could be used to assess the robustness of the protocols and equipment used in the Clayton Tunnel scenario, perhaps to give some insight into the pressures under which employees such as Killick worked. The user’s objective would be to perform Killick’s specified protocol, with a view to sending as many trains safely through the tunnel as rapidly as possible subject to progressively more frequent signal failure and ever more frequently arriving trains. The insight gained in this scenario is qualitative in nature and could naturally lead to a closer examination and more faithful simulation of ergonomic issues, such as train visibility and the time needed to reset the signal or deploy the emergency flags.

An alternative simulation exercise might involve playing the role of a driver unfamiliar with the signalling protocol who is expected to infer the conventions from experience. Being able to infer the significance of the signal protocol depends on how often a hazardous encounter with a train occurs after the signal has been ignored on *proceed with caution*. This is a function of train frequency, engine failure, and what protocols other drivers in the model are observing. In this case, the skill to be learned is associated with apprehending dependency and recognising agency.

These illustrations strongly suggest that the development of intelligent systems cannot be divorced from the learning process. As Kirsh indicates [55], if learning is understood as “acquiring knowledge of a domain”, there is an issue over whether two creatures with slightly different physical attributes who had learnt the same task in behaviourally different ways could be said to have learnt the same thing. From an EM perspective, it is simplistic to separate abstract knowledge from the experimental contexts in which this knowledge is demonstrated. By way of illustration, it is certainly possible for two pianists to give performances that cannot be distinguished by a listener, but to have learnt the instrument in radically different ways. For instance, one may rely totally on a musical score, whilst the other cannot read a score in real-time, but plays fluently from memory. In this case, there are experiments to distinguish between the two performers, such as switching off the lights.

It may be that, from the ideal perspective to which logicism aspires, the development of safe railway systems should have been a process that involved immensely careful empirical analysis and formal specification leading directly to precise implementation. Railway history, even to the present day, makes this scenario seem entirely implausible. Systems such as the Clayton Tunnel telegraph were experiments that served to disclose the discrepancy between abstract and embodied signalling protocols, and the essential role that experience plays in (as far as possible) confining agents, observables and interactions to what has been preconceived.

The convergence of technology and operating conventions towards interaction in a closed-world environment is the recurrent theme behind the evolution of modern railways. Technology can both assist and subvert this process. Had Killick and Scott been able to communicate by mobile phone, the Clayton Tunnel disaster could have been averted. But as recent experience on British railways confirms, new technologies potentially introduce new hazards.

The ontological and epistemological stance of logicism is appropriate in those regions of experience where the empirical foundation for a closed-world model has been established. It is inappropriate prior to this point, and (as is well-recognised) insufficient to express experiential aspects of interaction that can be very significant. Paradoxically, the realm over which logicism rules is one in which in some respects intelligence is least taxed. The autonomy and initiative that Killick and Scott exercised to such tragic effect is precisely what had to be eliminated from the railway system in order to enhance safety.

4 The Implications of EM

Having discussed the character and significance of EM from a practical viewpoint, it remains to return to the broad agenda set out in the introduction. This section discusses how EM contributes towards three key objectives:

- giving a perspective on logicist and non-logicist approaches;
- providing a conceptual foundation for AI broader than logicism;
- providing a context for existing practice in “scruffy” AI.

There is, in particular, an important need to provide a more coherent framework from which to view the diverse applications of EM that have been studied, and to understand their significance and implications. To this end, this section draws on critiques of logicism from many different sources and perspectives.

4.1 Critical Perspectives on Logicism

There have been many criticisms of the logicist position. Where AI is concerned, the sources include Rodney Brooks [31, 32], Brian Smith [70, 71], Mark Turner [73] and Peter Naur [63]. Other relevant philosophical ideas are drawn from William James's ideas on Radical Empiricism, first collected for publication shortly after his death in 1910 [53], and from more contemporary work of Gooding [47] and Hirschheim *al.* [52] on methodological issues in science and information systems development respectively. These indicate that the controversy surrounding a logicist viewpoint is neither new, nor confined to AI and computer science. Gooding's analysis of Faraday's work is motivated by disclosing simplistic assumptions about the relationship between scientific theory and practical experiment. William James addressed similar issues in his attacks upon the rationalist viewpoint on experience. Hirschheim [52] is concerned with information system design as involving the development of social communication systems. This arguably places the design of such systems outside the paradigm for Computer Science proposed in Denning *al.* [41]. For instance, it raises issues such as shared meaning, and the management of ambiguity, inconsistencies and conflict in system specifications.

Common themes that arise in these writings include:

- the role and limitations of language;
- the importance of agency, and of societies of agents;
- the significance of artefacts and constructed models;
- perception and action, observation and experience;
- the importance of situated and empirical activities;
- the significance of metaphor and analogy;
- the relationship between private and public knowledge.

A particularly significant challenge is the development of an appropriate computational paradigm for AI. Brian Smith [70] has sought such a paradigm for some time, and identified many of its essential characteristics. An important theme in Smith's work is the connection between computability and physical realisability. This endorses the philosophical position of Brooks [31, 32], and points to the difficulty of carrying out AI research without constructing physical models. It also argues for an emphasis upon empirical elements such as perception, action, observation and experience. Metaphor, analogy and agency have an essential role to play here.

The issue of whether we can address the many different aspects of a non-logicist position represented in the work of Brooks, James, Smith and Hirschheim without compromising conceptual integrity is particularly problematic. This is

illustrated by the breadth of interpretations of the term ‘agent’ in these contexts: Brooks refers to robots and humans as agents, and to the layers of the subsumption architecture as resembling a family of competing agents such as Minsky describes in [61]; James discusses agents as pragmatically identified elements in a particular causal account of our experience [53]; Smith declares that “agents are what matter for semantical connections” [70]; Hirschheim [52] is concerned with systems analysis that embraces machines, organisms, social and psychic systems, each of which represents agency of different character, and with design activities in which many human agents and different viewpoints are involved.

A detailed account of the relationship between EM and the work of these authors is beyond the scope of the paper, and the main emphasis will be upon giving an integrated view of where seminal issues have been or potentially can be addressed by EM. For instance: computer programming for AI is discussed in the context of Smith’s Two Lessons in Logic in [7, 20]; some of the issues for information systems design raised by Hirschheim *al.* in [52] have been considered in [17, 18, 64], and some relating to concurrent engineering in [1]; the status of *constructed models* in knowledge representation, as examined by Naur in [63], is discussed in [9, 12]; the prospects for applying EM principles to meet Kent’s challenges for data modelling [54] are considered in [11, 45]; layers of intelligence serving a similar role to those in Brooks’s subsumption architecture are introduced in [14, 4]; the ideas of William James [53] and Gooding [47] on the relationship between theory and experiment are developed in [9, 10, 8].

4.2 A Framework for EM

EM has been represented in this paper as primarily a first-person activity. The process of identifying similarities between two experiences is essentially an activity for one mind. Psychologically, it makes good sense for the primitive cognitive elements to belong to the first-person, but the bias of philosophy has tended to be towards explaining cognition with reference to third-person primitives.

If it is appropriate to ground experience through the perception of dependency in the way that has been described in this paper, then a new ontology is needed for many conventional concepts. EM imputes greater autonomy to the isolated learner, provided only that they have interactive access to physical artefacts. This makes it possible to view familiar concepts, such as objective reality, conventional theories and language, from a new perspective. Empirical activity and the recognition of dependencies are the means to account for these sophisticated communal concepts from the primitive base of private experience. Such a development is outlined in [15].

The aim of this section is to elaborate on the implications of this reconstruction process in three respects. There are useful links to be explored with other research. The conceptual reorganisation supplies a framework in which to classify the many practical applications of interaction with artefacts in EM. The reconstruction offers a new perspective on fundamental concepts such as agent, metaphor and intelligence that are arguably difficult to formalise satisfactorily [58, 35].

EM: the first person perspective. Under some interpretations, Kant’s famous dictum: “sensation without conception is blind” might serve as a motto for the logicist. An appropriate motto for EM might be that of the anonymous little girl who, on being told—by a logicist, no doubt—to be sure of her meaning before she spoke, said: “How can I know what I think till I see what I say?” [65]. This epitomises the first-person variant of EM that has been described above: the dialogue between me and myself, in which understanding is construction followed by reconstruction in the light of experience of what I have constructed. First-person activities in EM have centred on interface development [22, 10] and conceptual design [2].

EM: the second person perspective. Other applications of EM are concerned with the projection of agency from first to second person. The essential principle behind this projection is that through experiment and perception of dependency I can identify families of observables (constituting a ‘you’) who can be construed as acting in ways congruent to my own. This congruence is represented in the same way that I represent my own experience with reference to a suitable physical artefact. The projection process is conceptually simplest where the ‘you’ is another person, but this not necessary. Many non-human entities can be construed as agents in this manner: experimental science relies on just such principles to construct instruments that make interactions we can only conceive perceptible in an empirically reliable and coherent fashion.

Once a system is construed as having two or more agents, there can be ambiguity about the viewpoint on system state. Where there are several LSD accounts from different viewpoints, the integrity of observables becomes an issue. To an external observer of a system, this can manifest in many forms of conflict between the actions of agents within a system. It can also be associated with legitimate inconsistency, as in the case of Russell’s two observers, who see the same observables but have a different perception of the events. These two second-person variants of EM are represented in case-studies in concurrent systems modelling [16] and concurrent engineering [1] respectively.

EM: the third person perspective. One of the most complex and subtle processes that can operate in an EM framework is the transition to the third-person perspective. The observables that can be viewed from a third-person perspective are those elements of our experience that empirically appear to be common to all other human agents, subject to what is deemed to be the norm (*cf.* the presumptions surrounding the Mensa problem above). The identification of such observables is associated with interaction between ourselves and other human agents in a common environment. Objectivity is empirically shaped concurrently by our private experience, and our experience of other people’s responses.

The extent to which objective third-person observables dominate our public agenda can obscure the sophistication of the social conventions they require. In matters such as observing the number of items in a collection, and confirming its objective status, complex protocols are involved: eating an item is not permitted,

interaction with the environment must be such that every item is observed and none is duplicated, co-operation and honest reporting of observation is needed to reach consensus.

Underpinning third-person observables are repeatable contexts for reliable interaction, and associated behaviours of different degrees of locality and sophistication. In this context, locality refers to the extent to which a pattern of activity embraces all the agents in an environment and constrains the meaningful modes of observation. Counting techniques provide examples of behaviours that are typically local in this sense - they involve few agents, and are applied in the context of otherwise uncircumscribed interaction. Conventional computer programming typically presumes a closely circumscribed context, in which human-computer interaction is subject to global behavioural constraints (as in sequential interaction between a single user and computer), and the main preoccupation is with objective changes of state (such as are represented in reliable computer operation and universally accepted conventions for interpretation of input-output state).

Relating EM to conventional computer-based modelling. In principle, EM can be the prelude to the construction of a reactive system and more specialised forms of computer programming [13]. To understand the significance of using EM in this fashion, it is essential to appreciate that modelling behaviours is not a primitive concept in EM. In EM terms, a behaviour is a sophisticated abstraction that involves the attribution of an identity to a pattern of state transitions. For instance, a train passing through the Clayton Tunnel is associated with a sequence of observations that have some perceived reliability and integrity. In the context of Figure 1, the accident investigator will need to exercise all manner of changes of state of mind that are unrelated to the characteristic motion of the train. Observing a train in motion through the tunnel is choosing to generate a particular kind of experience from the artefact, one in which observation of position, velocity, acceleration and a clock are all involved. What is more, it is only these abstract observables that are significant: the identity of the train and all the particulars of the agents are immaterial.

In a classical approach to computational modelling of behaviour, the computer implementation begins from the premise that the empirical activities associated with circumscribing agency and identifying a closed world have already taken place. For instance, elementary physics declares how positions, velocities and acceleration are to be measured, and specifies the laws of motion that govern them. The construction of a computer simulation then involves:

- choosing a mathematical representation of the observables;
- specifying a computational model for the states and transitions;
- attaching a visualisation to display the behaviour to the computer user.

In constructing a closed-world model using the EM approach, these three issues are addressed in reverse order. The modelling process begins with the construction of an artefact whose primary function is to supply a visual metaphor for experience of its referent. Interaction with this artefact has an open-ended and

experiential character, but leads to the identification of stable patterns of interaction involving specific observables and circumscribed agency. Within the computational framework appropriate for EM, the modeller has unrestricted interaction with this artefact, so that its closed-world quality is respected only subject to discretion on the part of the modeller. It is possible to realise the closed world behaviour using a conventional computational model however. At this stage, a formal mathematical representation for the appropriate observables has been developed.

The development of a closed-world model is only possible subject to being able to construe a phenomenon in terms of patterns of interaction that can be entirely preconceived. In such a model, there is no uncircumscribed agency. This makes optimisations possible. For example, when we know that we are concerned with a train *travelling* into the tunnel, rather than having to consider the possibility of a train being *placed* in the tunnel, it is possible to specify that the train becomes invisible as it enters the tunnel. This is computationally much more efficient than maintaining a dependency between the visibility of the train and its abstract location. From an EM perspective, this change of representation is significant, as the model is then specific to particular contexts. In particular, certain “what if?” scenarios are excluded.

Multi-agent modelling in EM. Typical applications combine EM activity in first, second and third-person modes. This subsection reviews such applications within a multi-agent framework.

In the EM framework, artefacts are most appropriately classified according to the number of agents they involve. Informally, each agent represents a source of potential state change that cannot be—or is not as yet—circumscribed. By this criterion, the modeller is an archetypal agent. The variety of computational abstractions and applications is reflected in the number and nature of agents involved and whether their activity is or is not circumscribed. In a concurrent engineering framework [1], or a commercial spreadsheet-based programming environment [62], there are many modeller agents. In programming a reactive system [50], there are many agents whose interactions are yet to be circumscribed. Traditional sequential programming can conveniently be regarded as involving three agents: the programmer (uncircumscribed), the user (to be circumscribed) and the computer (circumscribed). Conventional formal systems are embedded within this framework as closed-world models in which there no uncircumscribed agents. Requirements analysis and formal specification of systems are processes by which we pass from agent-oriented models with uncircumscribed elements to circumscribed models.

EM typically generates a family of views of a concurrent system [1]. In many contexts, the modeller seeks a particular view—that of the objective observer who has a comprehensive insight into the global behaviour of a system. Such a view can only be developed in general (if indeed it can be developed at all) by a very complex process of evolution in which empirical knowledge has a crucial role. In principle, EM makes it possible to represent such an evolution by distinguish-

ing between asserting *what is observed* and asserting *what is believed* (cf. [2]). The computational forum for this representation is provided by the ADM [21], in which the modeller can prescribe the privileges of agents and retain total discretion over how these privileges are exercised. The evolution process converges if and when the modeller has specified a set of agents, privileges and criteria for reliability of agent response that realise the observed or intended behaviour. System implementation is then represented in this framework as replacement of certain agents in the model by appropriate physical devices.

More generally, EM can be applied in a concurrent engineering context [1], where independent views may be subject to conflict, as in Gruber's shadow box experiment [47]. To account for the process by which such views might be reconciled through arbitration and management requires a hierarchical model for agent interaction in which an agent at one level acts in the role of the human modeller in relation to those at the level below [1]. The associated "dialectic of form and process" is specified in terms of commitments on the part of the modeller agents similar in character to those involved in system implementation. Our investigation of a concurrent framework for EM of this nature remains at an early stage, but has direct relevance to requirements analysis [18] and has been used to devise simulations of insect behaviour illustrating Minsky's Society of Mind paradigm [56].

4.3 EM and the World of Pure Experience

The philosophical writings with perhaps the greatest relevance for first-person EM can be found in the work of William James. The connection between his agenda and that of the Empirical Modelling Project is clear from this preamble to "A World of Pure Experience" [53]:

It is difficult not to notice a curious unrest in the philosophic atmosphere of the time, a loosening of the old landmarks, a softening of oppositions, a mutual borrowing from one another on the part of systems anciently closed, and an interest in new suggestion, however vague, as if the one thing sure were the inadequacy of the extant school-solutions. The dissatisfaction with these seems due for the most part to a feeling that they are too abstract and academic. Life is confused and superabundant, and what the younger generation appears to crave is more of the temperament of life in its philosophy, even though it were at some cost of logical rigor and of formal purity.

James's philosophic attitude of Radical Empiricism has a yet more intimate affinity with EM. In some respects, this is best explained with reference to the status of an EM artefact as embodying interaction that cannot be expressed in a formal language. There is a practical communication difficulty familiar to all who have tried to describe the experience of interacting with definitive scripts. To make mischievous use of a blend that Turner would appreciate [74], it might be said that some of the most significant and subtle experiential aspects of EM interaction are nowhere more eloquently expressed than in James's essays.

Radical Empiricism and first person EM. In traditional empiricism, the content of what is empirically given is characterised as ‘discrete sensory particulars’ [53]. On this basis, driver Scott’s perception of danger on seeing the red flag cannot be viewed as a given. By taking this view, traditional empiricists found themselves in very much the same context as the modern logicist: they sought to derive fundamental features of commonsense experience from discrete sensory particulars by analytical processes. Bird [25], in his philosophical commentary on James, summarises James’s criticism of this stance: “Though the aim of empiricists was to identify the humblest, most basic, particular elements of the content of our experience they were driven in this way to deploy resources which were of a highly abstract, theoretical kind.”

In Radical Empiricism, James takes the view that the given primitives have to be recognised as typically more than discrete sensory inputs. In particular, though traditional empiricism recognises disjunctions as empirically given, and acknowledges separations between elements of our experience, it does not respect the significant conjunctive relations that also pervade experience: “the relations between things, conjunctive as well as disjunctive, are just as much matters of experience, neither more nor less so, than the things themselves”. Amongst these conjunctive relations, James cites identities, continuous transition, and “the most intimate relation . . . the relation experienced between terms that form states of mind”.

These ideas are quite coherent when transposed from James’s commentary on Pure Experience to first-person interaction with a definitive script. Variables in a script represent observables with identities. The changes to which their values are subject through interaction illustrate continuous transition in James’s sense (“this is the same observable, but, in the new context in which I am now contemplating it, it has another value”). New observables that are introduced into a particular script are recruited to the same state of mind (“I have observed—in addition—that Killick has a red shirt”).

James’s views on the primitive status of identities are particularly interesting in the context of the history of variables in mathematics (*cf.* [19]). For James, “conception disintegrates experience utterly”. It is in this spirit that the arithmetisation of geometry in the nineteenth century subverted the idea of a variable as a changing quantity. The historical importance of the relationship between dependency and the old-fashioned variable is highlighted by the Russian historian Medvedev [60]: “In this mechanical picture of the world the essential, one might even say definitive event was the concept of a law as a dependence between variable quantities”. (The translator’s pun was presumably unintentional!)

The open-ended and indiscriminate way in which new observables and different viewpoints can be adjoined to a definitive script distinguishes EM from conventional computer-based modelling in a striking fashion. It has its direct counterpart in experience in the way in which it is possible to embellish so freely on the observation of any particular state.

Two quotations serve to highlight the conceptual problems that confront the logicist who insists on interpreting observations as if they were elementary propo-

sitions. The first quotation is from James's contemporary F. H. Bradley [27], a strong rationalist opponent of James's views:

... mere experience ... furnishes no consistent view. [The direct products of experience] I find that my intellect rejects because they contradict themselves. They offer a complex of diversities conjoined in a way which it feels is not its way and which it can not repeat as its own ... For to be satisfied, my intellect must understand, and it can not understand a congeries [i.e. an aggregate] in the lump.

This quotation illustrates Bradley's disposition towards separating observations in an attempt to rationalise the world of sensations. Viewed from his perspective, conjunctions of such diversity appear to be contradictions.

James's response to Bradley is a rich source of reflections on how observations are experienced. It has incidental interest as a comment on the subject of 'consciousness', currently so topical in popular theories of AI:

To be 'conscious' means not simply to be, but to be reported, known, to have awareness of one's being added to that being ... The difficulty of understanding what happens here is ... not a logical difficulty: there is no contradiction involved. It is an ontological difficulty rather. Experiences come on an enormous scale, and if we take them all together, they come in a chaos of incommensurable relations that we can not straighten out. We have to abstract different groups of them, and handle these separately if we are to talk of them at all. But how the experiences ever get themselves made, or why their characters and relations are just such as appear, we can not begin to understand.

James's characterisation of Bradley's difficulties as ontological rather than logical accords with the account of definitive and logical variables above. In our experience of manipulating and interpreting definitive scripts of several hundred variables, definitions can seem to be chaotically organised, and do indeed typically combine observables that are incommensurable (e.g. colours, lengths, strings booleans). There is no natural ordering for the definitions in a script, and it is typically necessary to abstract different groups to interpret them effectively. This process of abstraction serves both to organise the script (e.g. to separate the graphical elements in the screen display into windows, lines and text), and to create experimental environments in which to study particular features of the script and its referent in isolation (e.g. to examine the operation of the telegraph outside the particular context of Figure 1). New features such as dependencies certainly emerge in the EM process through empirical investigation, but EM does not consider the underlying activity in the human brain.

Other prominent themes in EM are also represented in James. As Naur has also remarked [63], James pioneered the experiential view of knowledge, and recognised the importance of identifying concepts in psychological terms.

At the root of much of James's reflection is the idea of associating experiences within one mind. For him, knowledge is first apprehended as "[one] experience that knows another". The idea of 'bringing together experiences in one

mind' has much in common with the concept of blending that has been explored by Turner and others [74]. The ontological stance that James and Turner adopt is consonant with EM: the foundations of intelligence are to be sought in first-person experience, not in third-person abstractions. It is in this spirit that Turner regards the formal definitions of metaphor by logicians (*cf.* [35]), and the grammatical structure of a language, as sophisticated abstractions rather than primitive building blocks of human intellect. For Turner, the blend and the story (which to Bradley would doubtless have seemed so 'contradictory' in nature), are simple experientially-centred primitives.

Beyond the first-person. James's central focus in [53] is on a first-person perspective. Where a rationalist approach introduces the absolute as a common point of reference in discussing such personal matters as the perception of metaphor, James identifies a fundamental discontinuity "when I seek to make a transition from an experience of my own to one of yours ... I have to get on and off again, to pass from the thing lived to another thing only conceived" (*cf.* the strong distinction between 'I' and 'you' observed in Japanese, where certain verbs, such as forms of 'to want', can only be used in the first-person [51]).

In EM, the discontinuities between first, second and third person are in the first instance bridged by interaction with artefacts. All this is mediated by my own experience, even though this experience is of congruences between experiences that are first private, then of you, me and artefacts, and then of patterns that can be construed as reliably executed protocols involving many of us together with our artefacts (*cf.* [15]).

James's essays do not develop the ontology of second and third person in these terms to the full extent that is needed to account for EM modelling activities as a whole. His views nevertheless seem to endorse the perspective that EM commends. Central to his position is a pragmatic stance on the classification of experience: "subjectivity and objectivity are affairs not of what an experience is aboriginally made of, but of its classification".

As remarked above, many semantic categories of action are expressed by redefinition in a definitive script, but the distinction between them is potentially a matter of interpretation. It is possible that a feature first introduced into an artefact to improve the visual effect subsequently acquires a meaning in the referent. For instance, the point at which the engine triggers the treadle to reset the signal may be introduced as an abstract concept and subsequently be associated with a physical mechanism. The history of experimental science illustrates a similar recategorisation of experience [48]. For instance, in early microscopy, it was difficult to distinguish between optical imperfections and features of the object under examination. Likewise, in studying the Earth's magnetic field, there was at one time controversy over whether the Earth's field changed over time, or whether earlier measurements were inaccurate. These aspects of the process of construal go beyond the challenge to the adequacy of a logicist theory cited by Kirsh in [55]: *viz.* that it should be stated relative to "the actual subject-independent

properties of the domain". From an EM perspective, such properties can only be defined with reference to experiential criteria.

In EM, the computational setting of the ADM can be seen as a forum within which reality and agency are pragmatically shaped. In a discussion that is directly relevant to agency in EM, James considers how "the real facts of activity" should be construed. To paraphrase his account, he questions whether agency resides in: "a consciousness of wider time-span than ours", or "'ideas' struggling with one another, [so that] prevalence of one set of them is the action", or "nerve-cells [so that] the resultant motor discharges are the acts achieved". In arbitrating between these hypotheses, he advocates pragmatism: "no philosophical knowledge of the general nature and constitution of tendencies, or of the relation of larger to smaller ones, can help us to predict which of all the various competing tendencies that interest us in this universe are likeliest to prevail".

Drawing on previous research in EM, it is easy to envisage an environment for many designers concurrently interacting with a 'virtual prototype' for a complex reactive system (*cf.* [2]). Each designer has a personal library of scripts, each of which is more or less loosely associated with modelling a particular aspect of the virtual prototype. Corporate actions are decided through arbitration and mediated through dependencies and circumscribed 'intelligent' agents. In this scenario, the status of the virtual prototype is similar to that of an objective reality in Radical Empiricism. The agendas, conceptions and perceptions of the individual designers are represented via their private scripts, but can only be conceived by the external observer as an incomprehensible and incoherently distributed resource to be organised pragmatically in myriad ways for construction and experiment.

Such an image is consistent with James's description of his World of Pure Experience [53]:

Taken as it does appear, our universe is to a large extent chaotic. No one single type of connection runs through all the experiences that compose it. . . .space-relations fail to connect minds . . . Causes and purposes obtain only among special series of facts. The self-relation seems extremely limited and does not link two different selves together. On the face of it, if you should liken the universe of absolute idealism to an aquarium, a crystal globe in which goldfish are swimming, you would have to compare the empiricist universe to something more like one of those dried human heads with which the Dyaks of Borneo deck their lodges. The skull forms a solid nucleus; but innumerable feathers, leaves, strings, beads, and loose appendages of every description float and dangle from it, and, save that they terminate in it, seem to have nothing to do with one other. Even so my experiences and yours float and dangle, terminating, it is true, in a nucleus of common perception, but for the most part out of sight and irrelevant and unimaginable to one another.

In the light of this quotation, it is more understandable that Bradley deemed the products of raw experience to be contradictory. For him, as a rationalist:

“Truth . . . must be assumed ‘consistent’. Immediate experience has to be broken into subjects and qualities, terms and relations, to be understood as truth at all.” Inconsistency between the viewpoints of different designers may be explicable in logicist terms, but—as the above extract reveals—James also anticipated Minsky’s concept of conflicting agency within one mind [61]. Perhaps, as Kirsh remarks in [55], it is simply old-fashioned and parochial to hope for a logic-based denotational semantics for distributed AI systems.

4.4 Language from an EM Perspective

In *Experiment and the Making of Meaning*, Gooding writes:

Most received philosophies of science focus so exclusively on the literary world of representations that they cannot begin to address the philosophical problems arising from the interaction of these worlds: empirical access as a source of knowledge, meaning and reference, and, of course, realism.

Through its emphasis on first-person interaction with artefacts, EM is well-placed to provide an experiential perspective complementary to the literary world of representations. To meet Gooding’s challenge, it needs to do more than this: to explain the profound significance of language from within the framework of EM.

Language and learning in EM. The transition from first-person to third-person world in EM is associated with the empiricist perspective on learning set out in Box 3 (see [9]). As explained in [9], Box 3 is not intended to prescribe a learning process, but to indicate the nature of the empirical processes that construct a bridge from the realm of private experience to that of public knowledge. In Box 3, formal language is represented as relying upon the activities that are associated with the transition from first to second person, and from first to third. An informal account of how the use of language is related to other more primitive activities in Box 3 is given in [15]. This can be seen as parallel to the work of Turner [73] and Gooding [47] on what Quine terms “semantic ascent” [66].

A relevant perspective on language and computer-based modelling is given by Hirschheim *al.* [52], who consider the influence of philosophical paradigms on Information Systems Development (ISD). Following Burrell and Morgan [34], they identify four major paradigms that can be informally parametrised according to the stance they take on objectivism and subjectivism and on order and conflict. From the preceding discussion, it is apparent that EM does not respect this classification. That is to say, the products of EM can combine elements from any one of these quadrants, and EM activity can instigate transitions across any of these boundaries. In broad terms, however, concurrent engineering in EM originates in subjectivity and conflict (associated with *neohumanism*), and, like logicism, aspires to construct models that are ordered and objective (associated with *functionalism*).

private experience / empirical / experiential

interaction with artefacts: identification of persistent features and contexts
practical knowledge: correlations between artefacts, acquisition of skills
identification of dependencies and postulation of independent agency
identification of generic patterns of interaction and stimulus-response mechanisms
non-verbal communication through interaction in a common environment
phenomenological uses of language
identification of common experience and objective knowledge
symbolic representations and formal languages: public conventions for interpretation

public knowledge / theoretical / formal

Box 3. An Empiricist Perspective on Learning

Traditional ISD has been centred on the functionalist paradigm. In so far as logicism is essentially suited to specifying models that are ordered and objective, modern trends away from functionalism are non-logicist in spirit. Functionalism is charged with “treating the social world as if it were the natural world”. Where subjectivism is dominant, the primality of language is regarded as the only reality, whence reality is seen to be socially constructed.

In his construal of Faraday’s scientific work [47], Gooding is concerned to account for both the scientist’s interaction with others and with nature. For Gooding, an exclusively literary account of science can only be a caricature. In his view, modern philosophy of science “lacks a plausible theory of observation”.

The principles of EM can be interpreted as inverting the priorities of functionalism, so as to “treat the natural world as if it were the social world”. This cannot be seen as leading to socially constructed reality, since it gives interaction with artefacts such significance. In this respect, EM seems well-oriented towards Gooding’s requirements, denying the primacy of language, and perhaps also offering prospects as a theory of observation.

In claiming that language is not primitive, EM is in the tradition of the research of Brooks (“Intelligence without Representation” [32]), Turner [73], and James [53]. It is a claim that is widely challenged by philosophers. The contemporary philosopher Bird’s criticism of James expresses typical concerns [25]:

We are left . . . with a puzzle about the role or sense of ‘pure experience’. It is evidently of great importance in James’s account, and yet also totally inarticulate. . . .[cf.] Wittgenstein ‘a nothing would do as well as

something about which nothing can be said'. For James's pure experience has to be such that nothing can be said about it, if it is to fulfil the role for which it is cast. . . . Without some ability to characterise the experiences we have no means of determining their identity, and even no clear means of assessing James's central claim that we are presented with conjunctive relations in experience as well as atomic sensations.

EM is offered as a possible framework in which to address such concerns. The use of artefacts in representation can be seen as a variation on Brooks's theme of using the real world as its own model [32]. This is the device by which circumscription and commitment is avoided. Logician representations presume circumscription and commitment; after this presumption, a different ontology prevails. Such models can be derived from EM artefacts in the interests of efficiency and optimisation, but only at the cost of restricting to a closed-world functionality.

As Gooding remarks, agency—in the sense represented by first-person agency in EM—is absent from a scientific theory. To use Smith's words [70], mathematical modelling is *promiscuous* in character—the semantics of a model is not influenced by what formal abstractions are used in its specification, be they objects, relations, functions or agents. A plausible argument can be made that whatever *specific* behaviour can be observed in an EM artefact can be realised using a conventional program, and that on this basis there is no fundamental ontological distinction to be made. The most significant issue here is that, because of its status as a construal, an EM artefact resembles a single point in a space of referents within what the modeller deems to be a semantic neighbourhood. The problems of formalising unbounded determinism [42] are also relevant: the fact that the outcome of 'choosing a positive integer' can always be realised *post hoc* by 'choosing a positive integer not greater than a specified bound' does not mean that there is no distinction to be made between these two modes of choice.

The semantics of language from the EM perspective. Turner's thesis that "parable precedes grammar" [73] argues for an experiential framework within which to interpret language. EM has many advantages over logicism as a foundational framework for exploring this thesis. Making connections between the primitive elements of EM and informal uses of language is a more profitable activity than trying to express EM in formal terms.

The first, second and third person perspectives of EM provide different viewpoints on language. In first-person EM, words, as represented by the variables in a definitive script, serve as identifiers for observables. This resembles the interpretation of language that was initially proposed by Wittgenstein, and subsequently rejected in favour of a more sophisticated theory. In second-person EM, words figure as references to variables that represent observables of mutual interest to agents. The interpretation of such observables is negotiated in one of two ways: within a concurrent system through the shaping of communication protocols between agents, and in the interaction between concurrent designers through social convention. Third-person EM is the province to which formal language refers.

In studying natural language semantics, it seems appropriate to recognise that one word can be viewed from all three perspectives. In this way, there appear to be natural bifurcations in the meaning of certain words. From a first-person perspective, the word *time* can be used for *timely* (“now is the time for me to act”); from the third-person perspective, it refers to the objective *time as on a clock*. In the first-person, *real* means *authentically experienced*, and in the third-person *objectively existing*. In the first-person, *state* means *present immediate context* (“what a state I’m in!”), and in the third-person *abstract point in preconceived pattern of activity* (“once things get to this state, there’s no escape”).

From this viewpoint on language, it seems simplistic to assign discrete or narrow interpretations to words. Just as EM activity can migrate from personal, particular, and provisional worlds to the public, general and certain domain, so it seems can the meaning of a word. Turner’s account of metaphor [74] begins with a private experience that involves the blending of two spaces. To the logist [35], metaphor is expressed as a relationship between abstract structures that cannot necessarily be directly apprehended. To make psychological sense of the formal concept of metaphor, it is essential to trace its derivation from a private apprehension. EM supplies the appropriate context for this process.

The subtlety of the transformation of viewpoints that is characteristic of EM is not adequately represented by identifying first, second and third person perspectives. In developing EM in an engineering design context, it has been essential to stress the physical nature of the artefacts, and to invoke physical realisability in generalising classical foundations. This emphasis is consonant with Brooks’s concern for engaging with physical devices in AI, and with Smith’s thesis that “formality reduces to physical realisability” [70]. But this is not all that is required to represent our experience; as Turner has argued in [73], blending is a process that of its essence operates in the literary mind, and first engages with the products of our personal imagination. In accounting for individual insights and skills, there may be no clear objective physical points of reference. It is in this spirit that EM activity shifts independently along the private-public, particular-general, provisional-certain axes. The meaning of words can likewise migrate freely within this space.

The very concepts of learning and intelligence are profoundly connected with this process of migration. This accounts for the difficulty in formalising concepts such as *agent* [58] and *metaphor* [35] in a way that accords them the status of fundamental concepts of AI. An agent manifests first as an object-like collection of observables with integrity; then as an object that appears to be associated with characteristic potential changes of state. In some circumstances, the behaviour of an agent can apparently be so successfully circumscribed that its effect on state can be captured in a mathematical model. In this context, the concept of agent-as-object is hardly discriminating enough to be of interest, whilst that of ‘totally circumscribed agent’ forfeits the essential autonomy of agency. Yet, within an EM process, one and the same entity can migrate from one viewpoint to the other. What is more, in the migration process, it becomes essential to construe

the entity as an agent whose role can only be represented through recourse to first-person agency. To create the circumscribed closed world, it is essential to pass through the experimental realm.

5 Conclusion

Brooks has argued in [31, 32] that significant progress towards the principal goals of AI research—building intelligent systems and understanding intelligence—demands a fundamental shift of perspective that rules out what is commonly understood to be a hybrid logicist / non-logicist approach. This paper endorses this view, contending that logicism relies upon relating the empirical and the rational in a way that bars access to the primitive elements of experience that inform intelligence. EM suggests a broader philosophical framework within which theories are associated with circumscribed and reliably occurring patterns of experience. The empirical processes that lead towards the identification and formulation of such theories surely require human intelligence. The application of such theories, taken in isolation, is associated with rule-based activity as divorced from human intelligence as the execution of a computer program. Intelligence itself lives and operates in experience that eludes and transcends theory.

6 Acknowledgments

I am much indebted to all the contributors to the Empirical Modelling Project, and to Dominic Gehring, Theodora Polenta and Patrick Sun in particular, for their valuable philosophical, theoretical and practical input. Most of all, I am indebted to Steve Russ, whose constructive criticism and ideas have been crucial in identifying the essential character of EM. I also wish to thank Mike Luck for several useful references and feedback. The idea of relating EM to first-, second- and third-person perspectives owes much to several workshop participants, notably Joseph Goguen, Kerstin Dautenhahn and Chrystopher Nehaniv. I have also been much encouraged and influenced by Mark Turner's exciting ideas on blending and the roots of language. I am especially grateful to the Programme Committee and the Workshop sponsors for their generous invitation and financial support.

References

1. V. D. Adzhiev, W. M. Beynon, A. J. Cartwright, and Y. P. Yung. A computational model for multi-agent interaction in concurrent engineering. In *Proc. CEEDA '94*, pages 227–232. Bournemouth University, 1994.
2. V. D. Adzhiev, W. M. Beynon, A. J. Cartwright, and Y. P. Yung. A new computer-based tool for conceptual design. In *Proc. Workshop Computer Tools for Conceptual Design*. University of Lancaster, 1994.
3. V.D. Adzhiev and A. Rikhlinsky. The LSD engine. Technical report, Moscow Engineering Physics Institute, 1997.

4. J. A. Allderidge, W. M. Beynon, R. I. Cartwright, and Y. P. Yung. Enabling technologies for empirical modelling in graphics. Research Report 329, Department of Computer Science, University of Warwick, 1997.
5. M.D. Atkinson and et al. The object-oriented database manifesto. In *Proc Int Conf on Deductive and Object-Oriented Databases*, pages 40–57, 1989.
6. J. Backus. Can programming be liberated from the Von Neumann style? *Communications of the ACM*, 21(8):613–641, 1978.
7. W. M. Beynon. Programming principles for the semantics of the semantics of programs. Research Report 205, Department of Computer Science, University of Warwick, February 1992.
8. W. M. Beynon. Agent-oriented modelling and the explanation of behaviour. In *Proc. International Workshop Shape Modelling Parallelism, Interactivity and Applications*. University of Aizu, Japan, September 1994.
9. W. M. Beynon. Empirical modelling for educational technology. In *Proc Cognitive Technology '97, IEEE*, pages 54–68, 1997.
10. W. M. Beynon. Modelling state in mind and machine. Research Report 337, Department of Computer Science, University of Warwick, 1998.
11. W. M. Beynon, A. J. Cartwright, and Y. P. Yung. Databases from an agent-oriented perspective. Research Report 278, Department of Computer Science, University of Warwick, January 1994.
12. W. M. Beynon and R. I. Cartwright. Empirical modelling principles for cognitive artefacts. In *Proc. IEE Colloquium: Design Systems with Users in Mind: The Role of Cognitive Artefacts*, December 1995.
13. W. M. Beynon and R. I. Cartwright. Empirical modelling principles in application development for the disabled. In *Proc. IEE Colloquium Computers in the Service of Mankind: Helping the Disabled*, March 1997.
14. W. M. Beynon and M. S. Joy. Computer programming for noughts and crosses: New frontiers. In *Proc. PPIG94*, pages 27–37. Open University, January 1994.
15. W. M. Beynon, P. E. Ness, and S. Russ. Worlds before and beyond words. Research Report 331, Department of Computer Science, University of Warwick, 1995.
16. W. M. Beynon, M. T. Norris, R. A. Orr, and M. D. Slade. Definitive specification of concurrent systems. In *Proc. UKIT 1990, IEE Conference Publications 316*, pages 52–57, 1990.
17. W. M. Beynon, M.T. Norris, S.B. Russ, M.D. Slade, Y. P. Yung, and Y.W. Yung. Software construction using definitions: An illustrative example. Research Report 147, Department of Computer Science, University of Warwick, September 1989.
18. W. M. Beynon and S. Russ. Empirical modelling for requirements. Research Report 277, Department of Computer Science, University of Warwick, September 1994.
19. W. M. Beynon and S. B. Russ. Variables in mathematics and computer science. Research Report 141, Department of Computer Science, University of Warwick, 1989.
20. W. M. Beynon and S. B. Russ. The interpretation of states: a new foundation for computation? Technical report, University of Warwick, February 1992.
21. W. M. Beynon, M. D. Slade, and Y. W. Yung. Parallel computation in definitive models. In *Proc. CONPAR88*, pages 359–367, June 1988.
22. W. M. Beynon and Y. P. Yung. Definitive interfaces as a visualization mechanism. In *Proc. GI90*, pages 285–292, 1990.
23. W. M. Beynon and Y. W. Yung. Implementing a definitive notation for interactive graphics. In *New Trends in Computer Graphics*, pages 456–468. Springer-Verlag, 1988. also University of Warwick Computer Science Research Report 111.

24. W.M. Beynon. Definitive notations for interaction. In *Proc. HCI'85*. Cambridge University Press, 1985.
25. G. Bird. *William James*. Routledge and Kegan Paul, 1986.
26. G. Birtwistle and et al. *Simula Begin*. Chartwell-Bratt, 1979.
27. F.H. Bradley. *Appearance and Reality*. Oxford University Press, 9th edition, 1930.
28. P. Brödner. The two cultures in engineering. In *Skill, Technology and Enlightenment*, pages 249–260. Springer-Verlag, 1995.
29. F. P. Brooks. No silver bullet: Essence and accidents of software engineering. *IEEE Computer*, 20(4):10–19, 1987.
30. F. P. Brooks. *The Mythical Man-Month Revisited: Essays on Software Engineering*. Addison-Wesley, 1995.
31. R. A. Brooks. Intelligence without reason. In *Proc. IJCAI-91*, pages 569–595, 1991.
32. R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.
33. A. W. Brown. *Object-oriented Databases: Applications in Software Engineering*. McGraw-Hill, 1991.
34. G. Burrell and G. Morgan. *Sociological Paradigms and Organizational Analysis*. Heinemann, London, 1979.
35. J.A. Campbell and J. Wolstencroft. Structure and significance of analogical reasoning. *AI in Medicine*, 8(2):103–118, 1996.
36. K. M. Chandy and J. Misra. *Parallel Program Design: a Foundation*. Addison-Wesley, 1988.
37. E. F. Codd. The relational model for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.
38. J. Cohen and I. Stewart. *The Collapse of Chaos: Finding Simplicity in a Complex World*. Viking Penguin, 1994.
39. C. J. Date and H. Darwen. The third database manifesto. *Database Programming and Design*, 8(1), 1995.
40. S.V. Denneheuvel. *Constraint-solving on Database Systems: Design and Implementation of the Rule Language RL/1*. CWI Amsterdam, 1991.
41. P. Denning and et al. Computing as a discipline. *Communications of the ACM*, 32(1):9–23, 1989.
42. E.W. Dijkstra. *A Discipline of Programming*. Prentice Hall, 1976.
43. H.L. Dreyfus. *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT press, 1992.
44. K. Forbus, D.Gentner, A.B. Markman, and R.W. Ferguson. Analogy just looks like level perception: Why a domain-general approach to analogical mapping is right. *Journal of Experimental and Theoretical Artificial Intelligence*, 10(2):231–257, 1998.
45. D. K. Gehring, Y. P. Yung, R. I. Cartwright, W. M. Beynon, and A. J. Cartwright. Higher-order constructs for interactive graphics. In *Proc. Eurographics UK Chapter, 14th Annual Conference*, pages 179–192, 1996.
46. D. Gentner. Structure-mapping:a theoretical framework for analogy. *Cognitive Science*, 3:155–170, 1983.
47. D. Gooding. *Experiment and the Making of Meaning*. Kluwer, 1990.
48. I. Hacking. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, 1983.
49. D. Harel. On visual formalisms. *ACM Comms.*, pages 514 – 530, May 1988.
50. D. Harel. Biting the silver bullet: Towards a brighter future for software development. *IEEE Computer*, January 1992.

51. M. Hiraga. personal communication.
52. R. Hirschheim, H. K. Klein, and K. Lyytinen. *Information Systems Development and Data Modelling: Conceptual and Philosophical Foundations*. Cambridge University Press, 1995.
53. W. James. *Essays in Radical Empiricism*. Bison Books, 1996.
54. W. Kent. *Data and Reality*. North-Holland, 1978.
55. D. Kirsh. Foundations of AI: the big issues. *Artificial Intelligence*, 47:3–30, 1991.
56. N.S. Lam. Agent-oriented modelling and societies of agents. Master's thesis, Department of Computer Science, University of Warwick, September 1993.
57. D.B. Lenat and E.A. Feigenbaum. On the thresholds of knowledge. *Artificial Intelligence*, 47(1):185–250, 1991.
58. M. Luck and M. d'Inverno. A formal framework for agency and autonomy. In *Proc. 1st Inter. Conf. on Multi-Agent Systems*, pages 254–260. MIT Press, 1995.
59. D. McDermott. A critique of pure reason. *Comput Intell*, 3:151–160, 1987.
60. F. Medvedev. *Scenes from the History of Real Functions*, volume 7 of *Science Networks - Historical Studies*. Birhauser-Verlag, 1991.
61. M. Minsky. *The Society of Mind*. Picador, London, 1988.
62. B. Nardi. *A Small Matter of Programming: Perspectives on End User Computing*. MIT Press, 1993.
63. P. Naur. *Knowing and the Mystique of Logic and Rules*. Kluwer Academic Publishers, 1995.
64. P. E. Ness. *Creative Software Development — An Empirical Modelling Framework*. PhD thesis, Department of Computer Science, University of Warwick, September 1997.
65. A. Partington, editor. *the Oxford Dictionary of Quotations*. Oxford University Press, 1992.
66. W.V. Quine. *Word and Object*. MIT Press, 1960.
67. L.T.C. Rolt. *Red for Danger*. Pan Books, 4th edition, 1982.
68. J. Rumbaugh et al. *Object-Oriented Modeling and Design*. Prentice-Hall, 1991.
69. B. Russell. *The ABC of Relativity*. George Allen and Unwin, 1969.
70. B.C. Smith. Two lessons in logic. *Computer Intell. Vol. 3*, pages 214–218, 1987.
71. B.C. Smith. The owl and the electric encyclopaedia. *Artificial Intelligence*, 47:251–288, 1991.
72. M. Stonebraker and et al. The third generation database system manifesto. *ACM SIGMOD Record*, 19(3), 1990.
73. M. Turner. *The Literary Mind*. Oxford University Press, 1996.
74. M. Turner. Forging connections. In *this volume*, 1998.
75. P. Wegner. Why interaction is more powerful than algorithms. *Communications of the ACM*, 40(5):80–91, 1997.
76. D. West. Hermeneutic computer science. *Communications of the ACM*, 40(4):115–116, April 1996.
77. M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10(2):115–152, 1995.