Agent-oriented Modelling and the Explanation of Behaviour

Meurig Beynon
Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK
Invited Paper for the University of Aizu International Workshop on "Shape Modelling: Parallelism, lnteractivity and Applications", September 1994

Abstract

A new method of modelling, based on agents, observation and experiment, provides a framework in which to represent behaviour that is immediately experienced rather than circumscribed. Ths paper discusses the principles behind the method, and explores its relevance to topical controversies concerning explanation of the behaviour of complex systems.

1. Introduction

In a previous paper [2], we have argued that the formal concept of computation is inadequate for modern applications of computing. The result is a tension between principles and pragmatism that manifests itself in many areas of Computer Science, including

- abstract programming paradigms,
- complex software systems development,
- programming for AI and end-user applications,
- mathematical foundations of programming.

The essence of programming lies in prescribing and interpreting the behaviour of reliable state-changing devices [3]. Traditional mathematical modelling is oriented towards describing the behaviour of devices or systems that resemble conventional computers, whose reliability of operation can be taken on trust, and whose mode of interaction with its environment is preconceived and circumscribed. To meet such challenges as are encountered in programming reactive systems [4,10], reliable patterns of interaction between system components have to be identified, exploited and monitored throughout the programming process. For this purpose, we need rigorous computer-based modelling methods that take account of an evolving real-world situation and enable us to combine principles of abstract mathematical modelling with observation and experiment.

In this paper, we explore the limitations of classical mathematical approaches to describing, prescribing and explaining complex system behaviour, and outline our progress towards developing an alternative principled agent-oriented approach to modelling that can address these issues more effectively.

2. System behaviour: fundamental concepts

Few accounts of programming and specification issues do justice to the extraordinary subtlety and complexity of the concept of *system behaviour*. Conventional notions

about the nature of behaviour stem from two traditions: classical applied mathematics, where the behaviour of complex systems can sometimes be predicted by simple laws, and classical sequential programming (cf Harel's *1-person programs* [10]), where the mode of communication from machine to user, the mode of observation of machine by user, and the protocol for interaction between user and computer is entirely preconceived. Modern developments in applied mathematics and computing suggest the need for more powerful theories. The expressive power and utility of conventional mathematical models is one focus of attention in the controversy surrounding reductionism [6,8]. Formidable problems have been identified in the search for mathematical models for biological systems [6,7,8], and for reactive systems [10].

2.1. What is a behaviour?

There are many hidden presumptions behind the concept of a behaviour. To define a be-haviour we must choose what to observe, how we observe it, and when (e.g. how frequently) we observe it. An entity or attribute can only serve as an observable in a context where its integrity is respected in transformation. The behaviour is then defined by the pattern of changes of state to such observables. Special conventions for observation have to operate where the changes of state are continuous. In a concurrent system, we have to deal with simultaneous changes of state.

Informally, we can speak of a behaviour with reference to several observations of the same things at different times or on different occasions. There is an important distinction between behaviour as defined by changes of state that are immediately experienced (such as the behaviour of a group of people at a party) and the behaviour associated with preconceived reliable patterns of state change (such as the succession of the seasons). Traditional mathematical models are concerned with behaviour only in the latter sense, where the context for observation is circumscribed and what is to be observed is to some extent anticipated. For immediate experience of behaviour we have only to perceive identities, but recognising circumscribed behaviour requires some presumption about the reliability and repeatability of the observation.

The relationship between behaviour as immediately experienced, and behaviour as circumscribed is the central theme of this paper. Since circumscribed behaviour involves some preconception about what is to be observed it implicitly involves an expression of faith. Subject to invoking this act-of-faith, we can often use a mathematical model as a basis for description and prediction of a circumscribed behaviour - we know that Spring will follow Winter. Where behaviour is immediately experienced, in contrast, we are not constrained to focus our attention upon the pattern of change in *particular* observables. This lends an entirely different quality to our experience of a phenomenon, as there is no restriction upon the attributes that can be taken into account in observation. This theme is well-developed in the theological writings of Martin Buber. For instance, [5: p17]:

In the work of art realisation in one sense means loss of reality in another.

Everyday experience simultaneously involves circumscribed and immediately experienced behaviours. For instance, when we watch a play, only particular actions and attributes of the set and the actors are specified in the script of a play. From one production to another, and one performance to another, the precise activity on the

stage will vary in respect of unscripted details. In the circumscribed behaviour that the script defines, what is meant to be observed is declared by the playwright. What is actually perceived to happen by a theatregoer is a subjective matter - depending upon who does the observing. It is often important to be able to take account of both kinds of behaviour within the same model. For instance, we may need to redesign the layout for a control interface when, because of a coincidental resemblance outside the scope of our circumscribed model, a user has confused a clock reading with a dial.

2.2. Experiment

Experiment is a way of bridging the gap between immediately experienced and circumscribed bchaviour. In an experiment, the framework for observation of a phenomenon is set down: what observables are to be monitored, what events are to be recognised as significant. Conceiving an experiment presumes that is to possible to create the same context for observation many times. In these respects, experiment resembles circumscribed behaviour, but an authentic experiment also has elements of immediate experience. The experimentor typically has some autonomous control over the values of particular observables, and may have no clear expectation of what consequent effects will be obtained.

Experiment has a paradoxical quality. When an experiment is performed for the very first time, the outcome is a matter of immediate experience. When the same outcome has been obtained so many times that we are in a position to express faith in its reliability, we can reinterpret it as a circumscribed behaviour. The important and subtle point to note here is that the transition from behaviour immediately experienced to behaviour that is circumscribed is a shift in viewpoint on the part of the observer, and involves no change in the experimental activity itself. The faith of the experimentor is in no way a guarantee that the experiment will always fulfil expectations. From this perspective, it is useful to view all behaviour as immediately experienced, and circumscribed behaviour as behaviour immediately experienced but reinterpreted in the context of faith.

2.3. Explanatory modelling: the classical perspective

The question: What is an *explanation* of behaviour? has direct relevance for applications of computing in AI and engineering design. The classical scientific view, as posed and critically analysed in [6], is: an explanation of some phenomenon consists of deduction, from natural laws, of that phenomenon. By this definition, the idealised motion of a projectile is *explained* by Newton's Law. Certainly the invocation of Newton's Law is more than an explicit description of projectile motion - it expresses the way in which the expected behaviour of the projectile depends functionally upon the choice of initial parameters. We can also explain our definition of the trajectory with reference to experimental data showing that projectile motion is governed by Newtonian mechanics.

In interpreting the Newtonian explanation for projectile motion, it is significant that Newton's law is formulated relative to a particular observational context. We measure the mass and (x,y) position of the projectile, and observe a clock. We discount air-resistance, regard the projectile as a point-mass, assume constant gravity etc. By implication, other attributes of the projectile (e.g. its colour, its chemical consitution)

are presumed irrelevant, so that (say) the colour of a projectile does not explain its motion. In arriving at Newton's law, the law and the observational context are developed in conjunction through an experimental process.

One view of the scientific method presumes that every phenomenon can in principle be explained through a reductionist application of a hierarchy of natural laws. An ecosystem is explained in terms of organisms whose behaviour is explained in terms of proteins and macromolecules and DNA code ... until ultimately all behaviour is reduced to The Theory of Everything [8]. Cohen and Stewart [6] argue cogently that "[science] is a far less coherent structure than is admitted in the orthodox caricature of the scientific method", and that "we need a new kind of theory, in which suitable aspects of phenomenan can be understood without referring them to lower-level rules". One of their primary concerns is the intractability of the mathematical problems associated with analysing and reasoning about abstract behaviours associated with natural laws. In this paper, we argue that many of the most relevant issues cannot be addressed by conventional mathematical modelling - a radically different framework for modelling behaviour is required.

3. Modelling behaviour: the modern perspective

3.1. Models of behaviour for Computer Science

As discussed in detail in [2], classical mathematical models of behaviour provide an insufficient foundation for each of the key areas of Computer Science cited above.

In programming paradigms:

- incorporating user-interaction into preconceived patterns of behaviour is problematic,
- new media for communication force us to consider modes of behaviour that cannot be expressed formally in conventional terms,

In AI

- concurrent programming applications demand a higher degree of dynamism and mobility of processes.

In software engineering:

- abstract mathematical models of behaviour aren't well-adapted for incremental or evolutionary development e.g. in concurrent engineering, where refinement of the model occurs over a long period of time with reference to diverse and continuously developing modes of observation and experiment,
- abstract mathematical models of behaviour don't give sufficient insight into issues of engineering feasibility or into the potential implications of engineering decisions e.g. in respect of fault-tolerance,
- the design of an engineering product requires a different kind of justification from pure deduction from laws: there is a role for experimental evidence and for particular knowledge (e.g. rules about features).

and applications:

- logicist accounts of intelligent behaviour are ill-suited to dealing with e.g. the response of a robot to situations that are not preconceived,
- formal models of behaviour do not address the issues of associating form and content that arise when programming a robot,
- formal models of behaviour do not take account of the role of views in interaction in a concurrent system and in human interpretation of behaviour,
- the classical explanatory framework cannot be applied to qualitative concerns such as arise in aesthetic design or subjective user-interface preferences.

Computer methods have come to play an ever more significant role in the process of understanding complex systems. For instance, Cohen and Stewart [6] draw attention to the substitution of computer simulation for proofs. Without principled ways to carry out such simulations it is hard to establish the formal status of such an approach.

3.2. Insect colonies: describing and understanding behaviour

Insect colonies are often cited as a challenge for reductionist approaches to describing and explaining behaviour. They provide archetypal examples of the difficulties of reconciling specification of global system behaviour with prescription of component activities.

In [6], Cohen and Stewart cite Langton's Ant as an example of a cellular automaton with a simple rule-based specification that apparently always leads to a particular emergent behaviour. Langton's Ant illustrates that, even with explicit knowledge of the Theory of Everything for a simple system, it may not be feasible to deduce the properties of the system at a higher level of abstraction. By implication, the mathematical problems of accounting for the behaviour of colonies on the basis of the characteristics of individual insects are likely to prove at least as intractable.

Cohen and Stewart's concerns about reductionism reinforce the idea that the problems of understanding or developing complex systems demand much more holistic treatment than premature transformation to an abstract mathematical model permits. It may be that, in studying a real-world colony of insects, we are forced to acknowledge that the behaviour of individual insects stand in just such a relation to the behaviour of the colony as Langton's Ant exemplifies. Should we decide that this is the case, however, it must be on the basis of all available experimental evidence. To this end, it must be possible to take account of experimental input throughout the modelling process:

- to allow experimental input to influence the evolving model
- to enable the model to guide the choice of the experimental framework

To counter the idea that mathematical abstractions in themselves hold the key to the description and explanation of behaviour of complex systems, we develop a fictional scenario about the behaviour of colonies of imaginary insects, called ficts.

Ficts are one of the oldest forms of life on the planets Alias and Bias in the solar system of Elias. They are renowned for the extraordinary configurations into which

they organise themselves, and have been the object of scientific study on both planets. The principal and original inhabitants of Alias are the A-I: they have a highly intelligent and technologically advanced civilisation that has now colonised large areas of Bias. The inhabitants of Bias are the B-O: a people infamous for their powerful scent who are far less advanced in scientific matters, but have a very well-developed sense of humour.

The A-I have been studying ficts for many centuries. Crude depictions of fict configurations, or conficts, appear on the walls of caves from the first century C1 of the Alian calendar. Conficts were first depicted as religious icons in the tenth century, when they were typically carved on bark in idealised forms with a high degree of symmetry. For many centuries, identifying and interpreting geometric patterns in conficts was an important skill that could only be practised by Alian religious leaders. Recognising such patterns often required unusual powers of imagination, but anyone who questioned the objectivity of leaders, or attempted to give more accurate representations of conficts was regarded as a heretic, and risked persecution and death. One such heretic was the C15th Allan artist Leonilla, who developed much more refined methods of drawing, and was the first to construct realistic images of conficts that were later to be useful in scientific analysis. Since C21, the old religious taboos have been forgotten, as conficts have been represented by photographs, videos and computer simulations.

The theory that conficts consisted of colonies of ficts was well-established long before the invention of the microscope first made it possible to observe individual ficts. Microscopic examination made it apparent that conficts were of two kinds: active, when the individual ficts were in motion, and quiescent, when the ficts were at rest. Shortly after they discovered magnetism in C20, the A-I observed that quiescent conficts were arranged in orbits that followed the lines in the local magnetic field. On this basis, they were able to attribute the well-established 10 year cycle of confict patterns to the 10 year Elian cycle of solar activity and to explain the unusual behaviour of conficts in the vicinity of power cables. The organisation of active conficts at present remains a mystery to the A-I.

The B-O have a relatively naive scientific understanding of confict behaviour. As they are much smaller creatures than the A-I, and have more acute vision, they are able to see individual ficts directly, but -from the perspective of the A-I -are handicapped experimentally by the fact that active conficts are thrown into confusion by a B-O presence. For this reason, the B-O have only been able to study quiescent conficts in detail. Long before the A-I developed the microscope, the B-O already knew of the correlation between the clockwise / anticlockwise orientation of ficts in their orbits and the red / black colour of the ficts on Bias. The celebrated Bian scientist Bodkin tried to prove that the colour of ficts determined their orientation by dyeing black ficts red, but this experiment failed. This result would not have surprised the A-I, as - unknown to the B-O - there are red varieties of fict on Alias that adopt an anticlockwise orientation.

Neither the A-I nor the B-O have gained much insight into the way in which active conficts are organised. The B-O can only catch glimpses of active conficts, which degenerate into chaos as they approach. The B-O are foolish enough to suppose that ficts are very intelligent: they believe that ritual joke-telling ceremonies - similar to

those practised by the Bians themselves -account for the shape of active conficts, and that ficts are engulfed by the overpowering sense of humour of approaching Bian observers, and fall about laughing. The A-I, in contrast, fail to realise that ficts *are* very intelligent, and that in active conficts they organise themselves into patterns to leave scent traces that record the location of food sources, predators, local features etc in a complex language. Neither do they realise that the alignment and orientation to the magnetic field of quiescent conficts is a matter of social convention (dating from C13) that serves to minimise the energy consumption of the colony when it is not engaged in recording such information. It is the scent of the B-O that overpowers the pheronomes in an active confict - a fact quite inexplicable to the A-I, who have neither experienced nor developed the concept of scent.

4. Analysis of the fict fiction

4.1. Situated modelling

Our fictional scenario motivates a different perspective on the behaviour of complex systems from that suggested by focusing upon mathematical abstractions in isolation. Examination of the way in which the A-I and the B-O observe and explain the behaviour of conficts illustrates how our theory of confict behaviour will be influenced by

- what we choose to observe of the colony and the ficts
  e.g. Bodkin tries to correlate colour of ficts with fict orientation,
- what we believe ficts are capable of perceiving and affecting
  e.g. A-I don't appreciate that ficts communicate in forming active conficts,
- what we are ourselves capable of perceiving and affecting
  e.g. A-I can detect and perturb magnetic fields, and observe Elian solar activity,
- what we can accurately record / register in experiment with appropriate instrumentation / computation / perceptualisation
  e.g. A-I have no metaphor for scent, and no instruments to register scent,
- what conventions we establish for interpretation of what we experience
  e.g. A-I priesthood interpret conficts as religious icons.

Of course, our illustrative example is contrived, but these influences upon a theory of behaviour are all represented in carrying out a complex engineering task. In building an airbus, for instance, we have to isolate the significant observables and organise the activity of the human agents and the electronic components to operate within their performance capabilities. We base our construction and organisation upon experimental knowledge gained from simulating fragments of the system state, using appropriate metaphors and instrumentation to reflect state in a way that we can perceive and interpret. We rely upon conventions for the interpretation of experience both in the experimental process (e.g. training the engineers to interpret circuit diagrams and use oscilloscopes) and in the development of the final product (e.g. designing the instrument panel, training the pilot, taking account of aircraft safety conventions).

Analysis of the fict fiction leads us to appreciate the way in which the perceived behaviour of a system is a function of the characteristics and context for interaction of

the agents within the system and those of the observer. The modelling to which we aspire is aimed at articulating this dependency of system behaviour upon the characteristics of the constituents of the system and the perspective of the observer. The history of the study of conficts is intended to illustrate how radically any theory of behaviour depends upon the context in which it is conceived. This motivates models that can be developed and evolved in conjunction with experimentation in a real-world situation.

Activity of just this nature is represented in traditional engineering. The history of aircraft design illustrates the systematic refinement of a theory of behaviour, whereby knowledge is acquired through decades of experience of monitoring aircraft, pilots and passengers, of developing materials and technologies, of investigating aerodynamics in theory and practice. By comparing a modern aeroplane with its earliest ancestors, we can appreciate what a rich interaction between experimental insights and evolving theories has contributed to the development of flight. Over this period, we have accumulated what we presume to be reliable knowledge about what attributes of the aeroplane are relevant to its performance, how the choice of these parameters affects its behaviour, what conventions are appropriate for the cockpit layout and the pilot communication protocols, etc.

We use our fict scenario as a way of motivating "situated modelling" techniques [12], in which a model is developed in the context of a real-world situation. The models we can develop in this way are necessarily quite different in character from traditional mathematical models. The power of our development method derives from the idea of systematically refining our system model by using experiment to identify more and more precisely which factors serve to determine the behaviour. But if our theories about the behaviour of a system are open to refinement through experiment at any stage, they are also in principle subject to refutation. These observations point to a profound distinction between the two approaches to modelling.

4.2. Classical modelling

In making a mathematical model of the behaviour of a system, we create an abstract representation that can only be enriched by subsequent observation in preconceived ways. For instance, we may represent the behaviour by a differential equation of a particular form, and subsequently estimate the coefficients by experiment. In Langton's ant model, we make the presumption that everything we need to know about the ant is captured in the mathematical prescription of its behaviour. In such a context, it is meaningless to ask the question: can the ant remember whether it has visited a cell before?

A mathematical model of a system may be said to offer an *explanation* of a phenomenon in as much as the effects of changing a parameter can be predicted from theory. The quality of such an explanation depends entirely upon how precisely the actual system behaviour conforms to the mathematical model. There are many issues to be considered in assessing this conformance:

- to what extent is the mathematical model an idealisation?
  - the traditional objectives of behavioural abstraction are to guarantee freedom

from ambiguity and to eliminate conflicts, but an actual system typically behaves in a singular way in special circumstances;
- have we identified the only system parameters subject to change?
  - there may be many ways to view the behaviour of a system, and an abstract mathematical model typically only addresses one of these views;
- does the mathematical model relate the behaviour of a system to its structure?
  - the way in which a system is engineered profoundly affects its response to exceptional circumstances (consider e.g. the effect of a B-O presence upon conficts) and failure conditions (consider e.g. the implications of one fict acquiring the opposite magnetic polarity). An abstract mathematical model typically takes no account of this.

The quality of the insights that a mathematical model can give into a system depends upon how successfully the system behaviour has been circumscribed. Every invocation of the mathematical model is an implicit expression of faith in the reliability of the phenomena that inform its construction. The degree of faith that can be put in the abstraction is determined solely by the experimentation that precedes the conception of the model. The process of circumscribing behaviour abstracts patterns of agent activity that provide the basis for prediction. In this way, the autonomy of agents can no longer be expressed. This elimination of agents from a behavioural model can be seen as a virtue in some contexts (cf Hoare [11], Russell [13]) but in fact detracts greatly from its explanatory power. Further confirmation of the important role that metaphors for immediately experienced behaviour can play in complementing abstract mathematical models can be found in [9].

There are two principal respects in which traditional mathematical models are insufficient for explaining and developing complex systems. We need:

- a framework within which to express the modelling activity that precedes an expression of faith in the reliability of a phenomenon. Only in such a framework can we properly reflect the autonomy of agents, and correctly attribute features of the behaviour to their perceptions and capabilities.
- to be able to represent activities that involve circumscribing the behaviour of a system in ways that cannot be expressed formally in a conventional manner. For instance, the designer of an aeroplane will observe conventions in cockpit layout that have no basis in natural laws, but are nonetheless determined by the physical characteristics of pilots and the historical traditions that are implicitly represented in their training.

5. Agents, observation and experiments

5.1. Agent-oriented modelling over definitive representations of state: principles

The philosophical framework for studying behaviour considered in this paper is informed by a wide range of case studies associated with the development of new techniques of agent-oriented modelling. More detailed discussions of the relevant concepts and principles appear in a companion paper [15], and in other references (cf. [2]). In this context, we shall focus on identifying the features of our modelling method that are most relevant to the theme of describing and explaining the behaviour of complex systems.

In our modelling framework, observables are represented by variables. In observing a behaviour, there are certain indivisible relationships amongst observables, so that for example, the instant at which the minute-hand passes midnight may also be the instant at which a new year begins, and the point in time at which a savings policy matures. Such dependencies between observables are modelled by unidirectional constraints, and expressed using systems of definitions, or *definitive scripts*.

The semantics of a definitive script resembles that of a spreadsheet, in that it typically represents one state in a real-world behaviour that is immediately experienced. For instance, in a spreadsheet that represents the current state of a financial account, it is in general impossible to predict the form and effect of the next transaction. The use of definitive scripts as the fundamental method of representing perceived system state is consistent with the idea that behaviour that is directly experienced is more primitive than circumscribed behaviour (cf §2.2).

The power to model immediate experience distinguishes our approach from most other computational frameworks, many of which are based solely upon abstractions for describing circumscribed behaviour. Modelling immediate experience is the key to modelling experiment and to representing related activities, such as design. The typical interaction of the experimentor or designer is to modify a definitive script by redefining a variable. Such an action resembles "what if?" activity in a spreadsheet.

The concept of modelling immediate experience depends crucially upon the existence of appropriate metaphors. It is easy for me to convey the dimensions of the table upon which I am currently working, or to draw a picture to represent its shape, but I know of no way to communicate the smell of the dregs of tea in my mug with any degree of faithfulness. Significant factors in exploiting metaphors are the extent to which people can be trained to interpret metaphors (e.g. to understand a circuit diagram, or to interpret the visual display on a security system), and the extent to which the status of appropriate observables can be made perceptible. The power of the computer to animate metaphorical changes of state is the key to representing immediate experience. In animation, the choice of metaphor determines the values that are associated with variables in a definitive script, and the algebraic relationships that are used to express their dependency. The values of different types supply the underlying algebras for different *definitive notations* [2].

A definitive script can be seen as representing an agent's view of a complex system. We use the term *agent* to refer to an entity that plays a role in the representation and transformation of system state. Agents typically respond to changes in observables, and act through changing the values of observables. These activities are respectively modelled as monitoring the values of variables, and redefining variables. Simultaneous action on the part of two or more agents is generally possible, and is modelled by simultaneous redefinition. Agents are not necessarily persistent, and can be dynamically invoked or destroyed. The authentic value of each observable is bound to a particular agent, but can be directly observed and possibly changed by other agents.

In simulating the behaviour of a system of agents, the first step is the specification of those observables that are bound to an agent, those that is conditionally privileged to change and those that is able to observe. This specification is represented using the

special-purpose LSD notation. In animation from an LSD specification, we take account of the enabling conditions that must be satisfied before the values of observables can be changed, and the perceived events that serve as stimuli for agent action. The animation is executed in the computational framework of the Abstract Definitive Machine, an environment in which the user can act as a superagent to dynamically impose appropriate scenarios for action and interaction upon agents.

The degree of agent autonomy exercised in the simulation depends on the extent to which an ADM program is driven by human intervention rather than under preconceived automatic control. Where agent actions conflict, as for instance in an attempt to change the same observable simultaneously, this can be detected in the computational framework. In this context, the ADM user can act as an arbitrator to declare the outcome of such a conflict. This approach to modelling system behaviour is consistent with the idea that the concept of corporate behaviour of a family of autonomous agents is only meaningful with reference to an external mode of observation (cf §2.1).

To realise the full potential of modelling with agents and definitive scripts requires a more sophisticated machine model than the ADM currently provides. An important aspect of developing models that reflect experimental insight is the reinterpretation of immediately experienced behaviour as circumscribed behaviour via an act-of-faith (cf §2.2). One example of activity of this nature, familiar to experienced users of object-oriented development environments, occurs when generic features of different components are identified in a model, and these components can be replaced by instances of a new object. Another familiar process of abstraction arises in connection with the hierarchical decomposition of a model, such as is required to express the integrity of parallel activity of designers in the concurrent engineering process [1]. An extension of the ADM that supports these features is the focus for our current research, and is well-suited to our conception of explanation of system behaviour as described in §6.

5.2. Agent-oriented modelling over definitive representations of state: applications

The modelling methods that we are developing aim to capture the characteristic views and capabilities of all the agents in a system in an open-ended manner. Two complementary kinds of activity are involved in the development of our models:

- the refinement of agent models in the light of experiment;
- commitment to expressions of faith in how agents can be expected to operate.

The use of explicit state-based methods for representing agent views extends to the observers of the system, to include even the modeller herself. In this way - through the use of metaphor - it is possible to simulate experiment, and to represent the considerations that guide modelling decisions (such as the choice of an interface mechanism, or the construction of an aesthetically pleasing geometric object) for which there can be no formal specification.

Our philosophy of modelling leads us to identify a most significant distinction between the conception of system behaviour that is developed though real-world

observation, and the emulation of this behaviour by reliable state-changing devices. These two perspectives on phenomena correspond closely to top-down and bottom-up reductionism as discussed in [6]. Cohen and Stewart attribute the difficulty of marrying top-down and bottom-up analysis to the presence of a 'no man's land' they term Ant Country. They argue that the reductionist chain of logic does not traverse Ant Country, and that scientific explanation of complex phenomena accordingly relies not upon logic, but upon the use of analogy and "expressions of faith".

To justify our approach, we must go further in repudiating the reductionist position. For us, there is a fundamental qualitative difference between behaviour as observed in the real-world and behaviour as modelled using reliable devices. The gap between these two views of behaviour can only be bridged by experiment and expressions of faith. It is not simply that top-down analysis and bottom-up analysis "diverge into deductions too lengthy for the human mind to comprehend them" [6], but that there *is* in general no chain of logic to connect them.

To accommodate this shift in perspective, we have to give an alternative to the reductionist account of *explaining system behaviour*. Understanding through experiment how system behaviour depends upon the perceptions and capabilities of the agents and the mode of observation is the key to this. In the fict scenario, the orientation of ficts in conficts on Bias is perceived to depend on the colour of the ficts. This type of dependency may play a useful role in high-level reasoning about life on Bias, for instance, making it possible for an A-I scientist to determine the polarisation of a bar magnet by observing its effect upon a fict colony. But logically valid as this dependency is in its context, it cannot be interpreted as asserting that the orientation of ficts in a confict is explained by its colour, as Bodkin's experiment demonstrates. A more satisfactory explanation would be given by the discovery (say) that the presence of a particular gene was correlated with fict colour and fict orientation.

We can regard explaining behaviour as 'correctly' attributing system behaviour to characteristic properties of its constituents in accordance with the results of observation and experiment. By this criterion, the A-rs presumption that the organisation of quiescent conficts is the expression of a natural law is suspect. There is surely an objective criterion for preferring the explanation that quiescent conficts are organised by social convention rather than out of physical necessity. There will be certain experiments that can be performed to establish this, and perhaps historical evidence dating from pre-C13 to confirm.

6. Summary

The significance of an observation- and agent-oriented perspective to modelling can be summarised as follows:

it binds the form of the system model to its content.

Whereas the form of an abstract mathematical model of behaviour can be arbitrarily chosen so long as its interpretation stands in a preconceived relation to a real-world situation (cf. the *promiscuous modelling* discussed by Cantwell-Smith in [14]),

observation and experiment supply an objective criterion by which to discriminate between our agent-oriented models.

- it provides a framework within which to represent and systematically explore system behaviour that is imperfectly and incompletely understood.

  Traditional approaches to mathematical modelling of complex systems and the orthodox theory of computation promote the idea that the problems of understanding complex systems are essentially mathematical in nature. The fict scenario challenges the view that problems such as reasoning about the emergent behaviour of Langton's ant are properly representative of the issues involved in explaining system behaviour. For instance, Alian experimental knowledge and insight into active conficts is such that no mathematician, however ingenious, could account for their behaviour without developing entirely new concepts and abstractions. Our primary objective is to develop means to represent the partial information upon which theories of behaviour can later be established. The most important concern is that any abstract behavioural problems should be situated in an appropriate observational context.

- it offers an altemative to a fully reductionist account of system behaviour.

  Our explanations of system behaviour refer to autonomous actions of agents operating at many levels of abstraction, and to appropriate assumptions about the reliability of their response and the characteristics of the environment for their interaction. We do not see it as appropriate to explain the social conventions of the ficts in terms of activity at a lower level of abstraction, such as natural laws. In our view, the attribution of behaviour to agents is a crucial factor in providing models that can be used to investigate fault-tolerance, to frame experiments that can lead to refinement of the system and to adapt systems to new requirements.

7. Acknowledgements

I am indebted to Jim Viner and Valery Adzhiev for ideas that have motivated this paper, and to Steve Russ, Lee Suker and Simon Yung for helpful comments. I also wish to thank the Unversity of Aizu, the EPSRC and the Royal Society for their financial support.

8. References

[1] V D Adzhiev, W M Beynon, A J Cartwright, Y P Yung, A New Computer-Based Tool for Conceptual Design, Proc. Workshop Computer Tools for Conceptual Design, Univ. of Lancaster, 1994, 171 - 188
[2] W M Beynon, New Paths for Programming in Theory and Practice, Univ. of Warwick, Sept. 1992
[3] W M Beynon, S B Russ, The Interpretation of States: a New Foundation for Comptation, CS RR#207, Univ. of Warwick 1992
[4] F P Brooks, No Silver Bullet: Essence and Accidents of Software Engineering, Computer 20:4 (1987), 10-19

[5] M Buber (trans. R G Smith), *I and Thou*, T and T Clark, Edinburgh. 1966
[6] J Cohen, I Stewart, Why are there Simple Rules in a Complicated Universe?, Warwick Univ., March 1994
[7] J Cohen, I Stewart, *The Collapse of Chaos: Finding Simplicity in a Complex World*, Viking, NY 1994.
[8] R Dawkins, *The Blind Watchmaker*, Longman, London 1986
[9] R Feynman, R Leighton, M Sands, *The Feynman Lectures on Physics Vol. II*, Addison-Wesley, 1964
[10] D Harel, Biting the Silver Bullet: towards a brighter future for Software Development, IEEE Computer, 1992
[11] C A R Hoare, *Communicating Sequential Processes*, Prentice-Hall 1985
[12] P E Ness, W M Beynon, Y P Yung, Applying Agent-oriented Design to a Sailboat Simulation, Proc. ESDA 1994, Vol. 6, 1-8
[13] B Russell, *ABC of Relativity*, George Allen and Unwin, 1969
[14] B C Smith, Two Lessons of Logic, Comput. Intell. 3, 151-160, 1987
[15] Y P Yung, Agent-oriented Modelling using Observation and Experiment, also in this volume