Two Agent-Based Models of Trust in Social Networks

A thesis presented in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

Susan Elizabeth Street

2008

Abstract

Trust is a pervasive feature of human social interaction. Much of the recent interest in trust has been at the level of individuals and dyads. But trust is also important in networks, as it enables the formation and maintenance of social cooperation. Understanding this requires an understanding of how trust arises, functions, and is maintained within networks of people.

Developing understandings of how individual behaviours aggregate, and how they evolve within an environment that includes other individuals developing similar behaviours is a difficult task. One way that it may be approached is through computer simulation using agent-based models. This thesis describes the development of two agent-based models of trust.

Agent-based modelling is a novel method within the discipline of social psychology. The thesis first describes what agent-based modelling is, describes some of the situations in which it might be applicable, discusses how it might apply to modelling individuals in a social setting, and discusses the experience of developing the model.

The first model was based on a theoretical cognitive model of behaviour within a particular formal game that has been claimed to involve trust, the Investor Game. This model showed that a population in which all individuals are are pursuing similar optimal strategies does not generate any of the interesting behaviours that we would expect to see in real-world interactions involving trust and cooperation. This tends to suggest that modelling trust behaviours also requires modelling behaviours that are untrustworthy, and representing a full range of potential behaviours, including outliers.

The second model is based on a more naturalistic setting, on-line peer-to-peer trading through sites such as New Zealand's Trade Me, or eBay. In this model, individual traders carry characteristics that determine their reliability and honesty, and attempt to find effective strategies for identifying other traders' trustworthiness. This model suggests that, while providing traders with minimal guidance on strategies and allowing them to search for the best strategies may result in them finding effective strategies, this is not the only possible outcome. Somewhat surprisingly, effective trust strategies acted to contain unreliability, rather than dishonesty.

Acknowledgements

Over the years of writing this thesis has been a number of years in the writing, over which time a number of people have provided support in various forms. My first two supervisors, John Spicer and Stephen Hill, provided a good sounding board for a number of apparently unrelated ideas that I was juggling in the earlier stages of this work. The idea about modelling Trade Me came while Chelle Stevenson sat me down and explained why I should, or should not, trust various traders on Trade Me. I bought a couch for the post-grad room, and got a subject to model thrown in for free.

Over the last six months or so, a lot of people came to the fore in helping me to finish. It simply wouldn't have happened without Linda Jones and Ian Evans stepping in to make sure that that happened. I am enormously grateful to both of them for going out on various limbs for me. I suspect that quite a number of other people may have been engaged in making that happen, particularly Margaret Tennant, Jackie Koenders, the DRC, and an anonymous reviewer. Most of that went on in the background while I continued to plug on with the manuscript, so I am sure that there are others whose efforts on my behalf I don't know about, which makes it difficult to be able to acknowledge them. If this was you, thank you too.

Particular thanks are due to Ian Evans who stepped in very late in the piece to provide enormous practical supervision assistance to me in turning a pile of chapters, of varying qualities, into something much more presentable. His calm assurance that it was going to be entirely possible to bring it together into a complete thesis turned the tide for me, and it has made a huge difference to the finished thesis. Ian's encouragement, combined with Linda and Ruth Tarrant's simple insistance that I would get it completed, gave me the encouragement to make it so.

The practical support of a bunch of other people have been vital. My family have provided both the material, social, and emotional support to enable me to write the thesis. Mum provided much financial support, and a place to stay and work for the last few months that enabled getting the whole document into shape. Mike provided encouragement and material support, particularly in the early stages. Sandy provided a home and a family, and solid, sane, and calm support through a couple of years when things went from mixed up, to shattering, to downright terrifying. Neither she nor the kids had too much idea how diabolical a doctoral student can be in the final stages before submission. Now they get to meet me all over again when I'm not a doctoral student any more.

Last, but not least, this research was supported by a Massey University Doctoral Scholarship, for which I am very grateful.

Foreword

In a previous life, prior to beginning to study psychology, I was an engineer as a designer and consultant in electrical power systems. When I first entered the study of psychology, it was pointed out to me that I'd probably find it a little different to engineering. That comment proved somewhat prescient. This thesis in many ways reflects a series of questions that struck engineer abroad in the social sciences.

Electrical engineering students spend an entire academic career understanding and manipulating systems that are composed of many elements. Entering psychology was jumping into a world that was largely dominated by the in depth understanding of single entities. Unlike engineering, in psychology understanding the individual wasn't simply an essential precursor to understanding the system.

Large as the difference in thinking in terms of individuals versus systems was, I found that the the most dramatic difference in thinking involved time. In fact, to an electrical engineer, time was all but missing in psychology. Almost universally, theory and analyses were entirely static. That may not have been entirely strange in itself, but the language that was being used to discuss phenomena drew frequently on words like increase, change, and intervention. Psychology, as an applied discipline, is largely concerned with bringing about change, but the thinking and analysis was in terms of static, that is, unchanging, phenomena.

My interest in trust grew out of an entirely different set of experiences, this time as a somewhat absent-minded foreign student in Indonesia. Talking to Indonesian people, I was struck at how low their expectations of the trustworthiness of their compatriots was. In part that was understandable, as the country is plagued with endemic corruption, and petty crime like pick-pocketing is common. But at a more personal level, I had the frequent experience of people returning my wallet when I had left it in local stores. Even more strikingly, I had left my ATM card in an ATM, with the PIN number punched in. Someone found it, and came across the road to the mall in search of me. I wondered how such low levels of generalised trust had become entrenched, when individual people had shown an extraordinary degree of honesty.

From the two puzzles came this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis investigates trust located within a social network, through the development of two agent-based models, one cognitive and theoretical, and one based on more general psychological ideas applied to a naturalistic setting. Agent-based modelling is a form of simulation modelling. Simulation modelling, in turn, involves representing real world entities and phenomena in software. In agent-based modelling a number of individuals are represented using agents: code segments in a computer programme. An agent-based model consists of these agents, along with a further code segment that represents relevant features of the environment.

The first of the two models was based on the strategies that individuals apply to a particular task involving trust that arises in a formal game called the Investor Game. These strategies had been identified by previous researchers, using a combination of simulation modelling and experiment (Rieskamp, 2001). I was curious to see what would happen when the strategies identified were fitted back into a model, and in particular whether any interesting patterns would result from combining a number of agents that were behaving in similar ways. Where all of the agents were using the same basic strategies, it was difficult to get agents to act sufficiently dishonestly that other agents would refuse to trade with them. This model suggested that while the strategies represented in the model were well supported, they were not sufficient to represent trust in real world situations without also representing untrustworthy agents.

The second model was of traders in an online trading market, such as New Zealand's Trade Me auction web site. Traders entered the market with individual ideas on how to use the information available to decide whether or not to trade with each other. As the trading develops, traders learn from their own trading experience, and adopt elements of strategy from other more successful traders. This model showed that, while markets *can* find optimal strategies by traders sharing strategy information, this is not necessarily the case. It also showed that the strategies evolved may be more sensitive to unreliable traders than to dishonest traders, despite the lower losses that flow on from trades that fail due to failures to communicate or to complete a trade.

Agent-based modelling offers a novel tool for researchers in social psychology. Rather than considering the distribution of characteristics and patterns in a population or, conversely, the history of a single individual, agent-based modelling allows the researcher to explore the interactions of a number of individuals. Within a simulation, agents remain discrete individuals, retaining and developing individual characteristics, and interacting with other agents. Both the history of their interactions, and the individual development of agent individual characteristics can be monitored by the modelling programme.

The process of carrying out research using agent-based modelling, indeed of carrying out computer simulation more generally, is different to the process when using other research methods. In part, this is because simulation can access different aspects of a social situation. For example, much research is concerned with the distribution of, and relationships between, some variables. We have a toolkit of ways to think about estimating and describing the distribution of a variable in a population, and for describing and estimating the relationships between these variables.

There are fewer options, however, if we want to explore how situations develop, and how change happens. The practical difficulties in collecting and analysing data over more than a small number of intervals makes it difficult to explore phenomena in which dynamics are important. Simulation is a method that allows us to represent variables, and to observe how they may change as a process unfolds in time.

A second difference between simulation and other methods is also important, although less obvious. Simulation forces us to take a perspective on social situations that is directed towards systems. Specifically, simulation forces us to develop theory about systems and the components of a system. This is quite different from much of psychology, where the individual is regarded as the source of the mechanisms driving behaviour.

While much of the focus in psychology is on individuals, systems are common in psychology. Groups, families, and organisations form some of the most readily identifiable systems in social psychology, and some phenomena are essentially systems phenomena. For example, a group phenomenon, group-think, can be seen as a system failure rather than being the failure of individual members of a group. Beyond the immediate social environment, other aspects of the environment can act as components in a system. Psychology certainly acknowledges the importance of the environment in shaping the individual, for example behavioural psychology identifies the modification of behaviour by the outcomes encountered in the environment. And social psychology is interested in ways that individuals attempt to modify their environment, particularly their social environment.

As agent-based models are based on the direct representation of individuals and their environments, the development of an agent-based model demands specific theory on how individuals function in the environments that they are likely to encounter. This includes theories on how individuals function in their interactions with each other, and in response to their environments. But theory is relatively rarely in the form that agent-based modellers need for developing such a model. The difference between the forms of theory available, and that demanded for simulation modelling is such that one of the strengths of agent-based modelling may be in supporting and encouraging the creative development of theoretical ideas, and may also also lead us to ask different questions in other research.

An illustration of this is provided by the developers of the game "The Sims", a agent-based simulation game. The writers wanted to build the game based on research findings, but found that there was little in reported psychological research that could be applied to their game characters (Harris, 2003). While there is much literature on relationships and social interaction, most was not directly useful in producing realistic individual agents, with realistic relationships and interactions.

Agent-based modelling may be suited to exploring particular features of social situations, but as with any method, agent-based modelling is not universally applicable. Rather, it is likely to be most useful in relation to research questions in which the effects of nonlinearity are prominent. Situations in which nonlinearity is encountered are very common: these arise when the combined effects of a number of people acting independently are involved, or when system states are developing or changing in time. Whilst such situations are very common in social psychology, they cannot be investigated using conventional statistical methods, except in very rare circumstances. Methods for dealing with nonlinearity tend to be somewhat exotic at best, and many are difficult enough to apply to tractable physical systems, much less social systems.

Objectives and Overview

This first chapter begins by outlining the structure of the thesis. Following this introductory chapter, the next three chapters of the thesis provide some background on the method that I used for developing the trust models; a form of simulation modelling called agent-based modelling.

Agent-based models have potential in modelling nonlinear social systems. While some of the dramatic features of chaotic systems are well known, it is less well known that all that is required to generate complex behaviour in dynamic systems are nonlinearities in the system or even just one of its components. Far from being exotic, nonlinearities are pervasive in social systems. For example, anywhere it is possible to do, or not do, something, or to make a choice between a number of possible options, a nonlinearity is introduced. In Chapter 2 I describe the features of nonlinear systems, and review some of means of exploring and understanding nonlinear systems.

Agent-based modelling is a form of computer simulation modelling and, as such, encounters issues in the philosophy of science about what meaning and use we might make of modelling generally, and of computer simulation in particular. In Chapter 3 I review some of these issues. These issues remain unsolved in the philosophy of science. While modelling and simulation are worthy of investigation in their own right, this is outside the scope of this thesis. Rather, the issues are flagged, and possible stances considered from the perspective of a modelling practitioner. While I do not attempt to enter these arguments, I do state a position about how this form of modelling might be located in the setting of psychological research.

Agent-based models are not familiar to most psychologists, and so in Chapter 4 I describe what agents and agent-based models are. I also describe how agent-based models can be realised in software, and review some of the software that is available for agent-based modelling.

In addition to questions about the role of simulation modelling in a general sense, there are more specific questions as to what agent-based modelling might mean when used in psychological research. In Chapter 5 I explore some of the issues associated with adopting agent-based models as research tools. Agent-based models are used with an underlying expectation that they can represent real-world systems sufficiently well to enable some understanding of them. In this chapter I consider firstly what real-world entities the individual agents within a model might represent, and secondly what the assembled model and its outputs might represent.

All of the foregoing prefaces the application of agent-based modelling in two models of trust in networks of a number of individuals. In Chapter 6 I present a brief review of literature surrounding trust, in support of the development of two models of trust in a small model population. The literature review first addresses of some of the ways that trust has been conceptualised across a number of disciplines. It goes on to describe some of the ways that trust has been investigated in previous work, primarily within the discipline of psychology. Finally I summarise some of the theoretical approaches that might be useful in developing agent-based models of trust.

In Chapters 7 and 8 I describe two agent-based models that I designed and programmed to provide illustrations of the application of agent-based modelling. The two models were based on quite different theoretical approaches to trust. For the first model, described in Chapter 7, I based the agent's decision-making in a trust task drawn from game theory directly on a cognitive model of decision-making about trust in this game. In this model, it proved very difficult to make the agents behave dishonestly and mistrustfully. The model produced a homogeneous population of largely static and cooperative agents. In the second model, described in Chapter 8, I used more general ideas about learning, and set the trust task in a more realistic environment, based on an online trading site.

In Chapter 9 I draw some conclusions about the results from the models. In this chapter I also return to reviewing my experiences of using agent-based modelling as a method in the light of my experiences of using the technique with these two different models.

Having completed this overview, the main body of my thesis begins with a summary of some features of systems that are difficult to address using conventional research methods, and conventional statistical analyses. While at a first glance it seems that these features are a scattered collection of unrelated characteristics, there is a common genesis underlying them; all are a consequence of nonlinearities in components of a system. In Chapter 2 I summarise and describe some of the features and characteristics of dynamic, complex, large nonlinear systems.

Chapter 2

Systems, dynamics, and emergence

The domain of psychology is huge, touching as it does on the entire range of human experience. As befits a discipline that has such a catchment, many different research methods are used, each with strengths in accessing different features. This thesis proposes agent-based modelling as a new method that may complement other research methods used in social psychology. Agent-based models will be introduced in Chapter 4, but first this chapter will outline some particular features that the method might address: those surrounding dynamics and systems.

To an electrical engineer beginning to study psychology, the apparent absence of thinking in terms of dynamics and systems from some branches of psychology was striking. The temporal dimension seemed to be missing almost entirely; only appearing by implication in developmental psychology. This was a little unnerving in a discipline that is often involved in intervening to generate change and in observing, understanding, and measuring change.

A second thing that seemed to be missing was that there did not seem to be a transition in analysis from knowledge at an individual level to knowledge about systems of individuals. This was particularly striking in the case of social psychology because of its explicit interest in peoples' activities in the presence of others. Psychology, a science oriented toward the individual, often seemed to become much less specific when considering the interactions of a number of people.

These initial impressions were, of course, unfair. There is a body of work addressing how social psychology might investigate social situations in which dynamics are important. A collection of methods and examples of dynamic systems in social psychology are described in Vallacher and Nowak's (1997) book, for example. Nevertheless, the opportunities for applying many of these dynamic analyses remain limited by practical issues, such as the need for long runs of data.

Similarly, there are bodies of work on many different social structures that involve a number of individuals, such as families, groups, and organisations. Nevertheless, while the interactions of a number of people are clearly important in these settings, the transition from thinking about individuals, through thinking about individuals in their immediate social environment, through thinking about individuals as components of larger social entities involves sharp changes in theoretical bases.

These two features - situations in which dynamics are important, and situations in which the actions of individuals may aggregate into systems of individuals - seem entirely unrelated. But this is not the case, dynamics and aggregation are linked: they are characteristic features of large systems, and they become particularly interesting in large systems of nonlinear elements.

One dictionary defines systems as "a complex whole; a set of connected things or parts; an organized body of material or immaterial things" (Allen, 1980). There are important elements within this definition. Systems consist of a number of parts, and are not merely a collection of parts. Rather, components are connected together in an organised and complex way. The result, the system, may be considered a whole.

Using this definition, there are a number of ways that individual people might be seen to be located in social systems. Identifiable social structures, such as families or organisations, can certainly be seen as systems in that they are organised bodies of people. But people are also located in a web of social connections, without any structure necessarily being identifiable, and that interconnected web can also be seen as a complex whole. Systems are ubiquitous in human social life, and therefore in the social sciences, with the potential for social systems to arise anywhere that relationships exist between a number of individuals. While psychology is often less interested in systems than in individuals, there are many circumstances when we cannot understand individuals independently of the context of systems in which they might be located.

This chapter will first describe some of the features of large nonlinear systems. Thinking about systems is not necessarily a natural process (Resnick, 1994), so some of the general approaches to thinking about and understanding systems are described. Two facets of the behaviour of large nonlinear systems are discussed: their behaviour in time, or their dynamics; and the implications for the aggregation of individual behaviours.

Large systems

Systems arise in many forms, electricity grid, biological cells, organs and organisms, city councils, and computer software are all systems. In social settings there are, similarly, many examples of systems. Families, markets, and crowds of people are all examples where a whole entity exists through the complex interconnections of a number of individuals. In the case of families and markets, the functioning of individuals may be organised, either formally or through informal cultural rules. Less obviously, crowds are also organised, not from without, but as a result of each person acting in their own interests, but within the constraints provided by their neighbours. For example, where a rapid evacuation of a large number of people passes through a constriction, like a doorway, the crowd may form into something approximating a semi-circular shape (Bonabeau, 2002). That shape has not been imposed on the crowd, but is a result of individual people moving within the constraints of building shape and neighbours' locations; it is an example of self-organisation within the system of the crowd.

In a system, there are at least two levels of aggregation: the individual; and the system. There are, correspondingly, at least two levels at which we might understand the behaviour of a system. One approach is to treat the system as a *black box*, treating it as if it is a single whole entity. Exploring the system then reduces to exploring the characteristics and behaviour of that entity. Bonabeau (2002) provides an example of an agent-based model of an internet service provider (ISP) market, in which both individual customers and the ISPs are modelled as agents. In this model, ISPs are modelled as having the capacity to generate new product ideas, but the internals of how innovation arises within the company is not specified or modelled.

Formal approaches to understanding systems have been relatively recent. Over the last century, analyses of systems allowed the understanding and control of machines and processes. Analyses of control systems proceed by characterising the system in terms of the process that links inputs and outputs, and by adding feedback elements to produce a fast and accurate shift to a target output (Power & Simpson, 1978). As is usual in engineering work, control systems theory is particularly concerned with establishing predictable and stable operation of the system. In practice, this meant that early systems analysis work was concerned with maintaining systems within the range of linear operation, so that their behaviour was amenable to mathematical analysis. The analyses involved in system design were abstract, and entirely general.

Beyond relatively simple linear systems, as size, dimensions, and nonlinearities are added the formal mathematical analysis of systems rapidly becomes difficult, then impossible. Nevertheless, the methods of investigating and understanding more complex systems have been informed by the experience of analysing inherently simpler linear systems.

Other methods have been developed for more complex systems. General systems theories (Bertalanffy, 1971; Checkland, 1999) suggest analysing complex systems in terms of inputs, the process that produces outputs from those inputs, how the difference between target and actual outputs are fed back to producing a correcting action, whether a system will converge across a range, or whether it might have a variety of potential outputs.

Three factors are important in the development of complex behaviours in systems. Two may be unsurprising. It is reasonable that we might expect a system to show complex behaviours when there are a large number of interacting individuals. Similarly, we might expect complex behaviours when a large number of variables influence the behaviour of the system. The third factor complicating the behaviour of large and complex systems may be less obvious. The presence of nonlinearity in system elements, even in just one of the elements in a system, can produce dramatic effects in the system's behaviour.

Very large systems are the rule, rather than the exception in social systems. Social systems consist a large number of individuals. These individuals are complex, they are affected by a number of variables, and they can generate a large number of possible behaviours. The connections between individuals may be much more dense than the elements in, say, a mechanical system. The topology of their connections is not Euclidean; the social links between individuals may include links between individuals that would, in the absence of that link, otherwise be quite distant. The conditions for the generation of complex system behaviours exist in social systems.

The understanding of large complex systems has been constrained by the difficulty in analysing them. Relatively recently, the availability of computer simulation has allowed access to the analysis of large and complex systems, and systems with nonlinear elements. These are otherwise inaccessible through traditional mathematical methods, such as the solution of differential equations. Simulation analyses have revealed that the behaviour of very large systems can be surprising (Gleick, 1988). They can generate outcomes that are not immediately obvious if the behaviour of individual components is extrapolated directly, including chaotic behaviours.

Nonlinear systems

Nonlinearity, as the name suggests, means that the relationship between two variables cannot be described as a straight line. But not only can the relationship not be described in terms of a linear equation, it cannot be transformed into a linear equation. In less mathematical terms, nonlinear characteristics exist everywhere there is a discontinuity. For example, a step function is a very common nonlinear relationship: below a certain input level, one behaviour might result, and above it another quite different behaviour is triggered. Examples of step functions are pervasive in human behaviour, for example people change schools, leave a job, have children, ask someone to dance, make decisions, and join and identify with new groups. Our lives are punctuated by frequent and abrupt changes.

So pervasive are nonlinearities in the human world it would seem that linear analyses may have quite a restricted application, but many initially nonlinear characteristics can be made linear. One technique, commonly used in statistical analyses, is to transform the characteristic so it becomes linear. Another is to constrain a system so that it works within the range in which all elements have linear characteristics. Any system with at least one nonlinear element is a nonlinear system, but while it can be linearised, or held within a linear range it is reasonable to use linear analyses to investigate the behaviour of the system.

Where this is not the case, and nonlinearities can come into play within large systems, a variety of phenomena can arise that are not accessible to linear analyses. These include the appearance of unusual dynamic behaviours, including the appearance of patterns from apparent chaos, and extreme sensitivity to initial conditions. If the system can reach a state where the nonlinearity in this one element comes into play, linear analyses may no longer be appropriate and other methods are needed.

Superposition

An important feature of nonlinear systems that has implications for psychological research is that the principle of superposition does not hold for nonlinear systems. The principle of superposition says that we can dismantle a linear system, and the set of its inputs to a system into simpler subsystems and single inputs. Then we can find the solution for each of these simplified subsystems and inputs, and simply add them together. It is simple to demonstrate superposition at work in a linear equation, for example $Y = 4X$ and an input of $X = 10$ can alternately be broken up into four inputs $X = 1 + 2 + 3 + 4$.

We get the same result whether we calculate

$X = 10 : \; Y = 4 \times 10 = 40$

or

$X_1 = 1 : \; Y_1 = 4 \times 1 = 4$

$X_2 = 2 : \; Y_2 = 4 \times 2 = 8$

$X_3 = 3 : \; Y_3 = 4 \times 3 = 12$

$X_4 = 4 : \; Y_4 = 4 \times 4 = 16$

$Y = 4 + 8 + 12 + 16 = 40$

In contrast, if the element has a nonlinear characteristic, for example a step function

$$F(x) \begin{cases} X \leq 2.5 : \; Y = 0 \\ X > 2.5 : \; Y = 2 \end{cases}$$

We get different results

$X = 10 : \; Y = 2$ and

$X_1 = 1 : \; Y_1 = 0$

$X_2 = 2 : \; Y_2 = 0$

$X_3 = 3 : \; Y_3 = 2$

$X_4 = 4 : \; Y_4 = 2$

$Y_1 + Y_2 + Y_3 + Y_4 = 4$

The principle of superposition seems somewhat obscure, but it permits an assumption that is very important in research. When we can assume that superposition applies, we can infer that different constructs and inputs can be isolated, and researched separately. The results of different inputs and effects can later be recombined to apply findings in other situations. This allows phenomena to be separated, and each researched and analysed independently. It allows findings to be applied to situations with different combinations of features. Superposition allows us to sum the effects of a linear combination of inputs, as is done in multiple regression. Along the same lines, the principle of superposition allows us to assume that individual elements can be connected together, producing an assembly or group that has characteristics that directly reflect the sum of the individuals and their interconnections.

The presence of nonlinearity means that the assumptions that are enabled by the principle of superposition cannot be made. A second consequence of an aggregation that engages nonlinear elements is that the dynamics that may be generated have some unusual features. The aggregate effects of the behaviours of a number of people cannot simply be added. Where a system does operate in a range that puts any individual element into nonlinearity it may produce effects in a population, including some that are qualitatively different to the characteristics of individuals in that population.

Levels and aggregation

Psychology is a discipline that addresses human behaviour at many different levels of aggregation, from cells, through individuals, to systems of people: whether families, groups, teams, tribes or any of a large number of social structures. Between these levels are regions of transition, where it is no longer useful to think in terms of the individual components, as these have merged into a larger system of these units. These larger systems have their own characteristic behaviours, that of the system itself.

From the foundation of thinking about dynamic systems, we can see that adding elements to a system corresponds to making incremental changes in the system. Where superposition does not apply, these incremental changes cannot be reduced to simple additions. Furthermore, these small changes to the system can eventually produce profound changes in the system behaviour. When this happens the system has developed its own behaviours, arising from the collection of individuals that comprise the system, but the behaviour of the system is distinctly different from the behaviour of the individuals in the system.

The entire sphere of interest in psychology includes not only discrete individuals, but also the behaviour of groups and individuals within the setting of groups. A complete understanding might be expected to bridge the transitions between individuals and group systems. This transition depends on understanding the patterns that can emerge as individuals form into social structures.

Emergence

The transition between individuals and the social systems and structures that may arise from their interactions has been a controversial area in the social sciences. Exploring the transition means having to explore the non-trivial patterns that arise from the interactions of a number of nonlinear agents. Non-trivial, in this sense, means that the outcomes may be quite different to what we might expect from summing the outcomes from each individual independently across the whole population. Some of the patterns arising from large nonlinear systems have characteristics that might be described as emergent.

Emergence is a controversial concept, not in the least because it is defined quite differently in different disciplines. One definition, from philosophy, is that emergent properties are "genuinely novel properties that are irreducible to, and neither predictable nor explainable in terms of, the properties of their constituents" (Kim, 1999). This account of emergence has a long history in philosophy, and Kim (1999) goes on to say that substantially this version is used in scientific writing. More recently, the availability of computing power has led to advances in the mathematics of large nonlinear systems. Previously, these were impenetrable to mathematical analysis (Franks, 1967, cited in von Bertalanffy (1971)). The application of computing power to large nonlinear systems has showed that patterns can appear from the apparently random noise that complex systems can generate (Gleick, 1988). These patterns are unpredictable outcomes, quite different in character to the characteristics of the components of the system, and have also been classed as emergent.

One strand in Kim's definition is that emergent properties are unpredictable. With some qualifications, nonlinear systems are consistent with this element defining emergent properties. There are two related sources of apparent unpredictability in nonlinear systems. Some unpredictability arises because large nonlinear systems are extremely dependent on the initial conditions, and small differences in the conditions holding either within the system, or at the environmental boundaries of the system can produce hugely different outcomes. A further source of apparent unpredictability is found in the complexity of movement that can be generated as the system tries to approach a final position that is itself moving in intricate patterns. Large nonlinear systems, despite appearing to behave randomly at times, are deterministic, but the smallest variation may be sufficient to trigger the appearance of a new attractor, with entirely different results.

A second strand relates to irreducibility. The behaviour of large nonlinear systems cannot be reduced to the behaviour of the individual parts. The composition of the system depends on all of the parts, and reducing the system's complexity by removing parts eliminates the system itself. The system that is left may bear little resemblance to the target system. Further, we cannot work backwards from what is known about the state of a nonlinear system to derive information about the constituent components and relationships that go to make up the system. The appearance of new, entirely different attractors means that the system can reach a particular state via a variety of possible paths, associated with different attractors and different histories.

Although complex systems can produce features that are irreducible and unpredictable, there is a point of difference between these features and emergence as defined above. The behaviours of large nonlinear are explainable in terms of the properties of the constituent components, although that explanation is not necessarily simple. If we hold that the definition of emergence requires that the property is not explainable in terms of the properties of its constituents, unexpected and unpredicted patterns arising from chaotic systems should not be classed as emergent, although they remain irreducible to these constituent properties, and unpredictable.

Much of the debate around emergence depends on differences in the definitions of what is meant by *irreducible* or *unexplainable*. For example, water is the classic example of emergent properties. The argument is that the properties of water are not reducible to a combination of the properties of hydrogen atoms and oxygen atoms. This argument applies only to the properties of these substances in their elemental form. If we include among the properties of hydrogen and oxygen their properties in an ionic form, then the properties of water are explainable in terms of the properties of hydrogen and oxygen ions, the bonds that can form between them, and the consequent effect on the geometry of the water molecule. The emergent properties of water can be explained in terms of the properties of the components, their assembly, and interrelationships.

Kim (1999) also notes that there is a second group of concepts associated with emergence: these require that an emergent has causal powers. In particular, an emergent can influence the behaviour of the elements from which it has emerged. This definition places a requirement that emergent structures and properties can have a downward causation, from emergent to the elements from which it emerges (Kim, 1999). This requirement raises some difficulties in how downward causation might occur, and how downward causation from the system might be differentiated from the aggregated influence of the individual parts that make up the system.

Another feature that has been required of an emergent is that it is stable (Elder-Vass, 2005). Again, this requirement rather depends on how we define stable. If stability is only required to be in terms of a pattern existing for long enough to be detectable, this is a reasonable criterion. If, on the other hand, a more permanent stability is prescribed, this clashes with some of the suggested sources of emergence, namely that emergence reflects a system that is governed by a strange attractor (Newman, 1996), or from a bifurcation from one type of attractor to another (Morcol, 2001). Neither possibility is conventionally stable in the sense that its stability reflects a permanent system state.

Emergence takes on a slightly different air in the social sciences, in which emergentists are pitted against individualists. Social science emergentists, arguing that social structure is an emergent quality and is irreducible to individuals, claim that emergent properties depend not only on existing individuals, but on the structure's history arising from the actions of previous individuals that are no longer in the population (O'Sullivan & Haklay, 2000; Elder-Vass, 2005). King (1999) points out that individualists make no claim that they exclude individuals that are no longer in the population (King, 1999), and suggests that, far from being two opposing camps, emergentists *are* individualist (King, 2007), as the social world as a network of individuals linked by social relationships.

Dynamics of nonlinear systems

Linear systems produce a limited range of dynamic responses. A linear system may be stable, unstable, or marginally stable. A stable linear system, when shifted a little away from the stable position, will always change so as to return it toward to that stable position. While a stable system moves towards a stable position, an unstable linear system will accelerate away

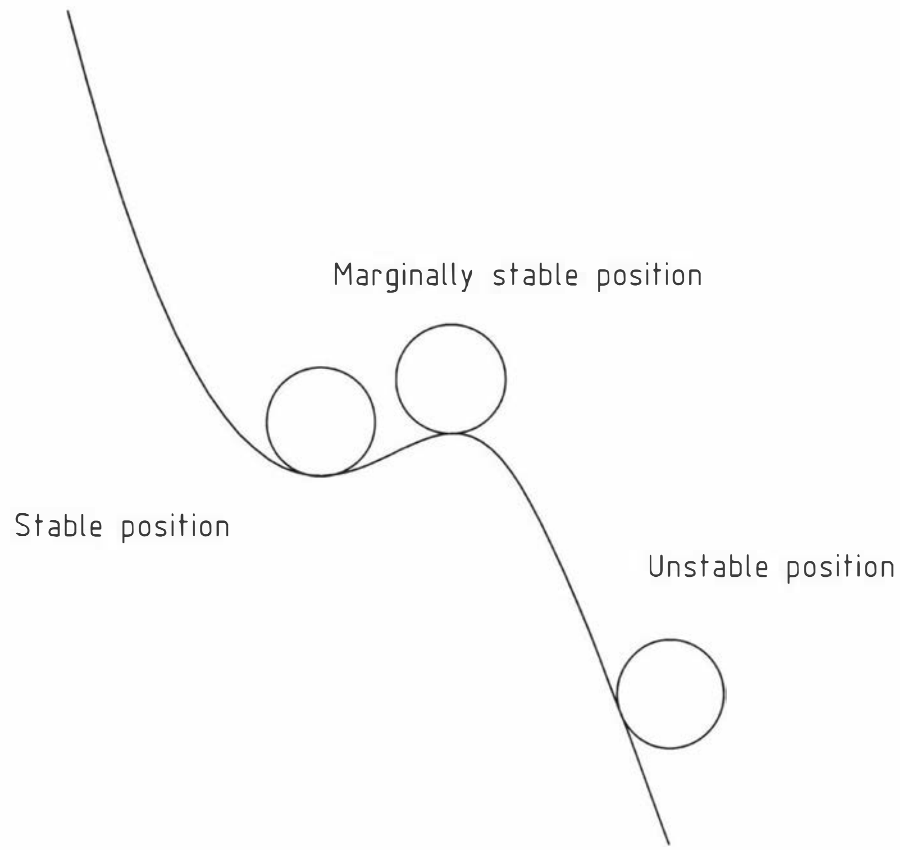

Figure 2.1: The possible dynamic outcomes of a linear system

from a corresponding unstable position. A marginally stable system can remain at a fixed position until it is perturbed. Depending on the perturbation, it might subsequently move either toward a stable point, or away from an unstable point. For example, this might be illustrated by the minimal system of a ball sitting on a slope (see Figure 2.1). In a stable position, the ball will not move, and if moved a little from that position, will return to the stable position. In an unstable position the ball would continue to roll downwards. In a marginally stable position, the ball would remain still until moved a little. Depending on the direction in which it is moved, it may either move to a stable position, or continue on an unstable trajectory.

In the previous paragraph, I have talked about a *stable position,* being the final state towards which a system moves. In nonlinear systems theory, these stable positions are called fixed point attractors (Nowak & Lewenstein, 1994). Mathematically, attractors are a set of limit values; theoretically the system perpetually heads towards this limit, but never actually reaches it. In practice, practical limitations, such as the precision of measurement eventually stop movement toward the limit, as the distance from

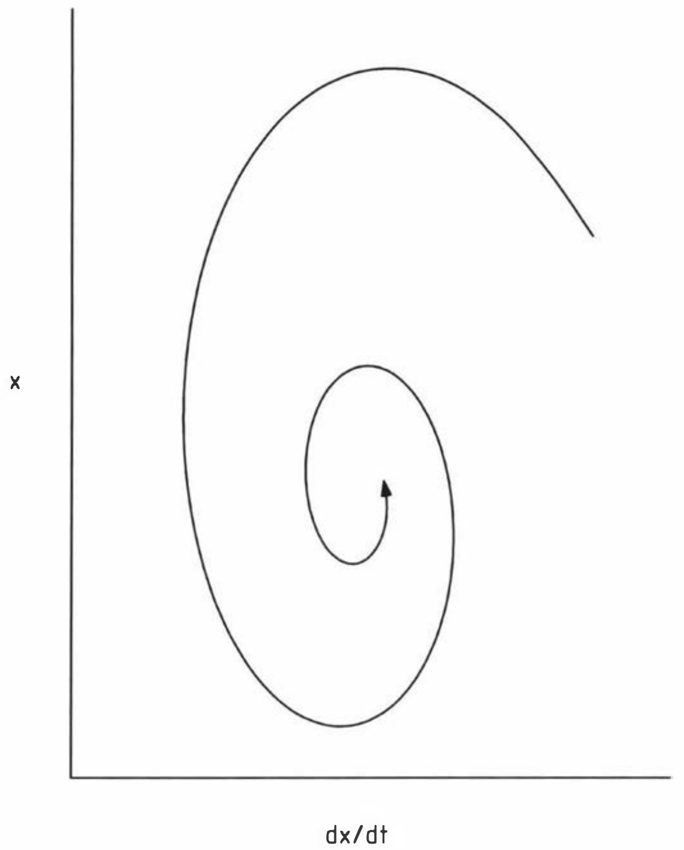

Figure 2.2: Phase space map of a dynamic system approaching a point attractor

the limit state becomes less than the measurement precision.

The behaviour of the system can be described through plotting the position of the system in time. While this does show the changes in the system as we might observe them over time, this plot does not necessarily make identifying the attractors easy. The system dynamics can also be described through a map of the phase space, in which the position is plotted on one axis, and the speed on the other axis. This figure does show the attractors. For example, Figure 2.2 shows a phase space plot of a point attractor. The corresponding motion of this system in time is shown in Figure 2.3. In this system, the system repeatedly overshoots the attractor. This appears as a damped oscillation in the position plot, and spiral path towards the attractor in the phase space map.

Fixed point attractors, such as are found in linear systems are only one of a number of possible attractors. While fixed point attractors are straightforward, attractors in nonlinear systems can be much more varied, producing a much wider range of potential outcomes. As well as fixed point

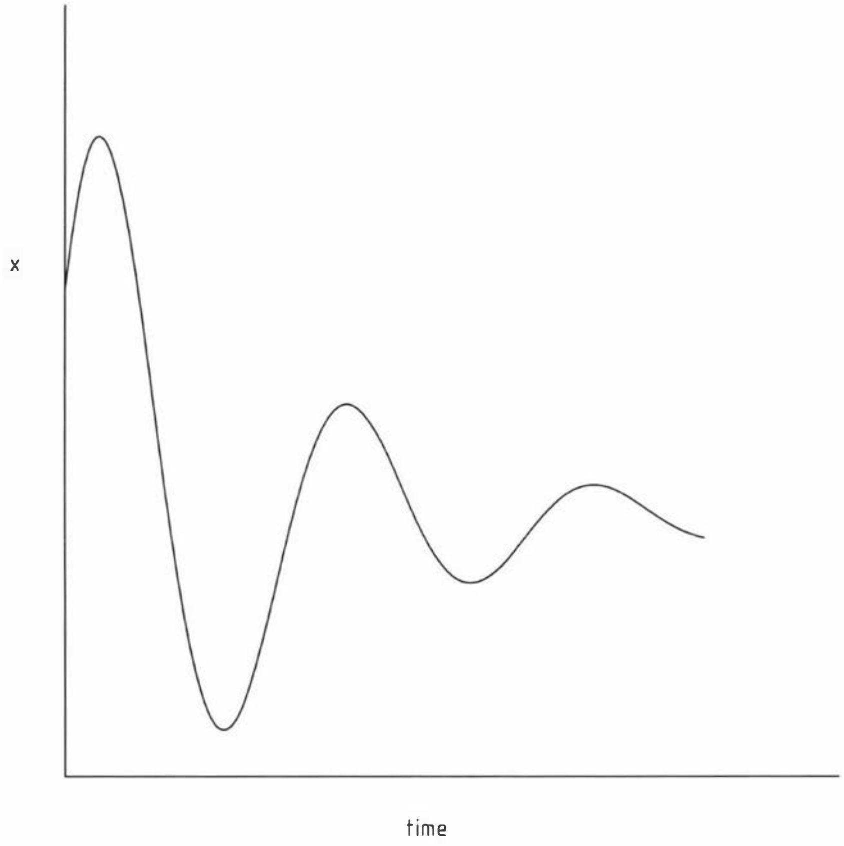

Figure 2.3: Position in time corresponding to the phase space plot in figure 2.2

attractors, nonlinear systems might have limit cycle attractors, multiperiodic and quasi-periodic attractors, or strange attractors, each of which has a characteristic geometry. The dynamics of the system are determined by the shape of the attractor, and by the path that the system takes in moving towards the attractor.

Limit cycle attractors appear in a phase space map as attractors with a closed loop shape. These result in the system following a periodic oscillation, for example the periodic patterns in population numbers as the fortunes of predators and prey alternate. Unlike a marginally stable system, a small perturbation from a limit cycle will be followed by a trajectory back toward the limit cycle (Nowak & Lewenstein, 1994).

Multiperiodic and quasi-periodic attractors appear in the phase space map as toroidal attractors (Nowak & Lewenstein, 1994). These attractors generate more complex periodic patterns than the limit cycle attractors. The underlying periodic components of the motion around multiperiodic and quasiperiodic attractors can be extracted using auto-correlation and Fourier analyses (Nowak & Lewenstein, 1994).

Strange attractors have very complex shapes, and may be fractals (Nowak & Lewenstein, 1994). Motion on a strange attractor can generate such complex behaviour in the system that it is difficult to differentiate from random noise (Nowak & Lewenstein, 1994).

These four forms of attractor can be found in nonlinear dynamic systems, and determine the types of movement that might be generated by these systems. While linear systems can generate a few types of motion, nonlinear systems can generate many different forms of motion. A further point of difference between linear and nonlinear systems is that, unlike linear systems, nonlinear systems can jump from one attractor to another. These shifts can result in the system suddenly changing the form of its movement, resulting in a sudden shift in the dynamic behaviour of the system. For example a system might suddenly shift from a fixed point attractor to a limit cycle.

It should be noted that, while the movement that they generate might be chaotic, large nonlinear systems are deterministic; their dynamics are entirely replicable if, and only if, we can replicate the exact conditions of the system and the environment. But very small changes in either the system, or in its inputs, can lead the system to an entirely different attractor. Thus nonlinear systems are very sensitive to small differences either in the system, or in the boundary conditions between the system and the environment. This sensitivity means that, although the system is deterministic, the behaviour of a nonlinear system is only repeatable if every element of the system and environment is identical. This can be done in computation, but is impossible to achieve in the real world.

A final common, and distinctive, feature of nonlinear systems is that they can exhibit hysteresis. Hysteresis means that the path in the forward direction, from A to B, differs from the path in the reverse direction from B to A. Linear systems, in contrast, are reversible. Hysteresis effects are widespread in social systems, for example, where a dyad traverses a relationship from formation to dissolution, the end position is not the same as the beginning position.

Time: the missing dimension

As I have said earlier in this chapter, entering the study of psychology from another discipline can be surprising. It is almost as if outside developmental psychology a whole dimension, time, is missing. This dimension is important in peoples' everyday lives, at least in the modern world. But despite this, time makes a relatively rare explicit appearance in the psychological literature. In contrast, in most disciplines outside the social sciences, it is almost taken for granted that time is an important dimension.

There are, of course, some good reasons for the low profile of time in social psychology. In a social science setting, it is particularly difficult to collect enough sequential data points to allow analysis of a time series. But this does not entirely explain the absence of time in other forms. For example, research that uses the rate of change of a variable as a construct is rare, although this is a more accessible variable as demonstrated by researchers who have used the rate of change toward a goal as a variable (Hsee & Abelson, 1991; Lawrence, Carver, & Scheier, 2002).

While it is relatively rarely used as an explicit variable, time does make a less explicit appearance. Any research that calls for observations made at a number of different times has an inherent time component, although this may not be acknowledged explicitly. Doing so does, however, make the assumption that any systems involved have reached their final stable state. Without some knowledge of the likely system dynamics, it is difficult to set the appropriate period that should elapse before a second set of observations is made. Different causal elements are likely to act over different time periods, a feature that Bandura notes works in our favour, as makes it possible to separate the elements of a causal network, for example allowing us to investigate the linkages in a triadic causation one by one (Bandura, 1986, p. 25).

There may be other reasons that explicit references to time might be missing in psychology. Much psychological research is carried out over a single time step, or a small number of intervals. Research in psychology rarely involves continuous observation of a process. That may act to make thinking in terms of time less automatic. Some who do incorporate time in theorising about psychology take quite radical views in theorising about time. For example, Levine (2003) has argued that the assumption that time is linear and constant may not be universally appropriate in social psychology. In part, his argument is based on the point that processes in psychology tend to occur in steps or cycles, rather than being continuous processes. In effect, he argues that we might deal with nonlinear characteristics by modifying how we think about time.

The physical sciences do have established techniques for modifying how we think about processes that have a time dimension. For example, we are accustomed to thinking about our 230V mains power supply in our homes as constant, rather than time dependent. But the notation $230V_{RMS}$ is actually the result of a transformation to remove time dependence in describing a sinusoidally alternating voltage that is time dependent, and that can be described mathematically by the equation $V = 325 \sin 100\pi t$. Techniques like this are available in the physical sciences because they have the luxury of components, whose behaviour is regular and relatively simple.

Levine (2003) does not propose ways that we might modify our conceptualisation of time in social psychology other than an arbitrary manner. In the sense in which he talks about thinking about time - in terms of steps or cycles - time becomes a more qualitative dimension.

These points about the representation and conceptualisation of time have a parallel in some forms of simulation, including agent-based modelling. Simulations using digital computers use iterated processes; the simulation is inherently carried out in steps. This can leave the simulation unscaled with respect to time, or at least unscaled with respect to linear, continuous time. However, if we are able to think about social processes in terms of steps and cycles, concerns about whether simulations should be time-scaled are possibly misplaced. A stepped simulation may be a good way to represent a process that proceeds in steps.

Others argue that incorporating time into our thinking in social psychology inexorably leads to thinking in terms of dynamic systems (Geert, 1997). There are many analogies that might be made between phenomena in social psychology and features of complex dynamic systems. For example, an alternative way to frame the sharp changes that Levine (2003) sees as challenging linear conceptions of time is to regard these as potential markers of nonlinear dynamic phenomena such as bifurcations or catastrophes. If we do so, the question in turn becomes one of how these ideas are applied, in particular, whether they are applied in a formal sense, or whether they are used as descriptive similes. Arguing for the former, Vallacher and Nowak (1997) maintain that while the early stages of use of new ideas might reasonably generate intuitive ideas about potential applications these need to be backed up by more formal approaches to using the ideas. Others have pointed out that attempts to do so have been relatively unsuccessful. One possible reason is that the thinking required does not sit comfortably with thinking about social phenomena (Puddifoot, 2000), although it is less than clear why thinking in terms of linear regressions is a more natural way of thinking in terms of behaviour. One possible hint as to the difficulty lies in his comments that the language used by advocates of dynamic systems thinking is in terms of physical systems, a criticism that could as validly be applied to my own explanation here. As Resnick (1994) has noted, systems thinking is not entirely natural, and the few visual similes that we have tend to be mechanical. A second part of the reason for this is that there is no social science equivalent of the language that mathematicians and physical scientists have developed for thinking and talking about systems and dynamics. The adoption of the language of mechanical systems has not been as successful for social scientists as it has, for example, by electrical engineers. As a result, the generality of thinking about dynamic systems is not yet as recognised as other general systems perspectives have become.

It is reasonable to say, as Puddifoot (2000) does that the comprehensive treatments of the analysis of complex nonlinear systems presented by proponents for dynamic systems approaches (Nowak & Lewenstein, 1994) looks difficult to implement. Further, depending as some of the methods do on long time series of observations, some of the techniques advocated suffer exactly the same problems that statistical time series analyses do. That is, it is difficult to collect the necessary data in social science settings, no matter what analysis you have in mind.

Nevertheless, there are approaches to thinking about phenomena in ways that do acknowledge system dynamics. One method that is underused in data analysis is the use of phase space plots. While these may seem more technical analytic tools (Puddifoot, 2000), they are actually an immensely simple tool for visualising data, including data that cannot be managed using formal analyses. For example, a short series of twelve measurements may be sufficient to produce an informative phase space plot.

As with the example of phase space plots, it may not be that the techniques needed for analysis of dynamic systems are any more esoteric than the statistical methods that are widely used in psychology. Rather, it may be a case of using different techniques. For example, where we are exploring how a system behaves, exploratory statistical techniques may be more appropriate than hypothesis-testing techniques. An example of this arises in the online trading model. With no obvious pattern in the results, I looked for groupings in the results using an exploratory technique - cluster analysis.

The dangers of thinking about systems in linear terms, and other than in systems terms, have been raised by a variety of writers (Bertalanffy, 1971; Checkland, 1999; Resnick, 1994). Equally, there is danger in thinking about processes that are located in time as if these were static. While some phenomena can be captured by what is effectively a single snapshot, or by spotting the differences in a pair of snapshots, not all can. There are a number of ways of thinking about systems that have been suggested by proponents of dynamic systems thinking. While some of the techniques offered may be difficult to apply directly to real world data, the generality of the approach means that the techniques can as validly be applied to data obtained from simulations. The alternatives are either to draw on the techniques of systems and dynamic thinking, or to try to capture the sense of processes that are in motion through, at most, a pair of static images.

Summary

This chapter has described and characterised some of the phenomena that we encounter in large nonlinear systems in general. The characteristics of social systems are such that these are large nonlinear systems with very complex elements. This being the case, collections of interacting individuals should be expected to produce phenomena that reflect this fundamental nature of the system.

Despite this, in many situations we can obtain good information and understanding treating a system as if it is a linear system. This includes good information about how individuals act within a system, and good information about particular phenomena.

There are, however, problems in applying the analytical techniques proposed for complex dynamic systems directly to every situation. There are times that systems effects, and dynamics are important, and the simplification of linear systems approaches cannot capture effects and phenomena.

Later in this thesis, I will apply these methods to two different settings involving trust. Trust is a phenomenon in which the mechanisms are described at an individual level, whether those mechanisms are thought to be cognitive, emotional, or personality characteristics. There are, however, features at a population level that seem to be related to the characteristics of individuals in the population. These will be detailed in Chapter 6. Relationships across levels like this involve aggregation of the effects generated by many individuals, and may need to be understood in this context. Attempts to make changes at a population level may need to make provision for the aggregated effects of individual level trust decisions and actions.

Chapter 3

Modelling and Simulation

Modelling

Before introducing a particular form of modelling, it is worth considering how and why we use modelling in science. When we model, the focus of interest is not the model itself. Rather, it is some real world situation - in the case of social psychology, some real world social phenomenon. These may be difficult to understand directly, because they may have any of a number of features that make understanding them directly difficult.

Why do we need to model?

One thing that can make real world social systems difficult to understand is their size. We may be able to track the activities of an individual, but as the number of interactions increases, the dimensions and complexity of any but the simplest social networks rapidly outstrip our ability to track and understand them.

Another difficulty that can arise for research in social psychology is the observability of the system, that is whether or not we can actually make the observations that we might need to be able to adequately understand what is happening. There may be difficulties in assembling complete information, even for small groups of people. The detail available from any one individual may be limited, and some individuals may not be observable at all because we may not have access to them. When we do have access to them, large scale measurement may be too expensive, or otherwise impracticable. Further, there are other constraints on what can be done in research. For example, it may be unethical or unsafe to collect data, or to manipulate a situation experimentally.

Even if a system is accessible enough to be observable, we can encounter features that make understanding the data conceptually difficult. In particular, large dynamic systems are difficult to understand. This difficulty is captured even in the vocabulary surrounding dynamic systems: large dynamic systems are also called *complex* systems, and their seemingly random behaviours are described as *chaotic*. Small and simple dynamic systems can be awkward enough, because we may have trouble amassing sufficient measurement intervals to permit time series analysis of a system that is changing or developing in time.

Where we cannot directly access, manipulate or comprehend a real-world situation through existing techniques, some other strategy is needed if we are develop our understanding of it. Modelling is one such strategy.

What is modelling, and how does it work?

One view of modelling is that it is a mapping process, in which a model is mapped to the real world (Holland, 1998). A model is a restricted likeness of a real world object or phenomenon. It represents important features of the real-world situation, but in a simplified way. A good model is accessible, both in that we can readily collect data from it, and in that its structure is easier to understand. One way to make a phenomenon more readily understood is to reduce the complexity. In modelling this means that not every detail of the target situation is mapped by the model: a good model includes representation of the elements essential to capturing the target phenomenon, and excludes the non-essential elements.

While all models involve simplification, different models vary in how they represent their targets. Scale models are one familiar form of model, that look like their targets in physical form and layout. But while representing the target visually, scale models do not represent their targets in other ways. They may be made from quite different materials, and there may be no attempt to represent any of the functionality of the target. For example, a scale model of the brain looks similar to a biological brain, and shows the relative location of major structures, but does not represent any brain functions. Scale models are not common in social systems, in part because they are difficult to set up, and in part because the characteristics of many social processes may change unpredictably as the number of individuals is increased.

The models used in social psychology are usually much more abstract in form than are scale models; many are mathematical models. These are notated in the form of mathematical symbols that do not bear the slightest resemblance to the individuals whom the model represents. For example, a model of children's television viewing might take the form:

$TVH = 0.25PS + 0.06HR + 0.11RT - 0.08V - 0.09VR - 0.66Co + 0.69(CoC) + 0.41(AP) + 0.68(AC) - 0.8(ACP)$ (Krosnick, Anand, & Hartl, 2003)

Where TVH is viewing hours, P is punishment style, HR is household rules, RT is non-school reading time, V is parental values for self-direction, VR is viewing rules, Co is coviewing, C is parent-child contact and A is age.

There is nothing in the form of this linear equation to suggest that it might represent children watching television. Indeed, we need a key to the variables to even get a hint that it might be a model of television viewing. Structurally similar equations could be used to model any of a variety of things, from building costs, to crop yields, to oxygen uptake in an athlete.

The process of statistical data analysis is one in which mathematical models are derived that optimise the degree of mapping between the model and data drawn from the real world. The scientific process does, however, involve more than the observation, mapping and replication of real world phenomena and data. It also involves interpretation of results, reflecting on these, and building theory (Haig, 2005). It generates a form of the model that is quite different to the form of the social phenomenon it represents, but, nevertheless, such a mathematical equation might be useful in describing some features of a population, and might be used to make predictions about individuals drawn from that population.

The process of statistical analysis has links with both theory and with data collected in the real world. Theory guides the form of the mathematical functions that we are going use in our statistical analyses, while data provide the points in the real world to which these are mapped. The role of models in the data analysis role, where data from the real world are mapped to mathematical functions is familiar and understood, but the role of models in building theory is possibly less familiar..

Models and theory

In saying that models may be useful in developing theory, there is an implication that models are distinct from theories. This is not necessarily a given. While the use of models is widespread, their role, meaning, and relationship to scientific theories is far from clear. Part of the difficulty is that models are described as being representative of the real world, or of theory, without any clear and universal understanding about what it is to be representative (Suarez, 2003). This is, however, not the only issue surrounding the use and status of models. Rather there are a number of philosophical questions: about what models are; what they mean; how we can learn from them; and how they relate to scientific theory (Frigg & Hartmann, 2006).

One influential stance, the semantic conception of theories, is that theory exists as a coherent set of mathematical models (Glennan, 2000). This set of models is either in the form of a set of formal logical or mathematical statements, or as a set of possible system states and transitions (Suppe, 1989, p.4). While this view proposes a relationship between theories and models, it only relates to a very specific form of model, a formal mathematical model of a very particular type (Suppe, 1989, pp. 39-41), or a set of possible system states. Models that consist a set of very formal, non-linguistic, mathematical or logical statements models might readily be seen as possibly having a role in defining a theory.

Importantly, the semantic conception is specifically a theory about theories, and not a theory about models. As such, it has nothing to say about how modelling is used in other ways in science. One form of model is regarded as a constituent component of theories (Glennan, 2000), but the semantic conception does not attempt to account for other types of model, or for other the roles of modelling in science beyond theory construction.

In the case of some models, the suggestion that they may not have a role in theory is unsurprising. For example, it would be a stretch to suggest that a Matchbox toy forms part of some theory. The Matchbox toy concept might go to suggesting that there are models that are not encompassed by the semantic conception of theories. Although the semantic approach is particularly concerned about models as components of theories, there are a number of linkages that can be made. Not only is there a link between models and theory, but also between models and real world phenomena, and between models and data (Frigg & Hartmann, 2006). Understanding how models are used in practice in science requires linking these into some coherent structure and process, but this requires a theory of models, rather than a theory of theories.

An alternative view is that models have a distinctly different role to theories. One version of this, the mediating models approach (Morrison & Morgan, 1999), suggests that models are entities in their own right, like tools or instruments. These entities are located between theories and the real world, and mediate between these (Morrison & Morgan, 1999, pp. 10-11) through a number of partial commonalities between the model and theory, and between the model and the real world.

Unlike the semantic conception of theory, this approach allows for a wide range of possible forms of model, and for a number of different models of the same phenomena. Models have a number of points of commonality, and different points of commonality may be useful in different circumstances, making it possible that different models can apply to the same real world phenomena. Under this approach, experiments might also be regarded as models of real world phenomena (Gooding & Addis, 2006).

One of the driving ideas of the mediating models approach is that a theory of models should recognise the way that models are used in practice in the sciences. In practice, models are dynamic entities, that are shaped as they are engaged by the scientists using them. Understanding comes from designing the model, and through manipulating and interacting with it.

Models in practice

There is little literature that specifies a process for building models. In practice, people building models use a mixture of devices: fragments of theory, mathematical tools, similes that draw on well-understood processes, assumptions and simplifications. The combination and assembly of these into a model is a creative, rather than a mechanical process.

The components, assumptions, and simplifications mean that any model is valid only over a restricted range. The model of television viewing above was derived with a population drawn from youth in the USA. One consequence of this restricted range of models is that even if a particular model is informative, we should expect the scope of its usefulness to be restricted.

As well as being developed from data derived from particular populations and phenomena, models are constructed with particular questions, or a particular class of questions, in mind. Within its design range, a good model should map its target well. There should, however be no expectation that the model should perform well outside its intended scope. For example, where we have developed a static model of a phenomenon, we should not use this model for approaching questions that may may be driven by an (unrepresented) dynamic process. This may seem obvious, but the meaning of a static model is often translated in terms like "for an increase in one unit of the IV there is an increase of two units in the DV". We tend to think of static models in dynamic terms.

As noted earlier in this section, psychology has an existing, well-established set of modelling techniques. For example, students are taught techniques for developing and stating theory using boxes representing constructs and arrows linking these. These, combined with the methods of statistical analysis, have been spectacularly successful in allowing psychology to become established as a scientific discipline. But having a constrained set of research tools comes at some cost. Each technique tends to be at its strongest when applied to particular forms of relationships in the data, and each model works well only within its bounds. With some thought, many research questions can be framed so that they are amenable to investigation through one of these techniques, but the other side of the coin is that the question asked often becomes determined by the available analytical tools.

In some cases, research questions remain unasked because we know that we cannot analyse them using standard techniques. These unaskable questions are not especially arcane. For example, we often test whether a property differs significantly between two different groups. We do not ask whether a property is the same for two groups, because we do not have suitable techniques for testing for this. The best we can do is to fail to find a difference. There is an asymmetry in how we frame questions. Having a broad and varied repertoire of research methods enables a wider range of questions to be asked.

An advantage of having solid and established techniques for theorising, model building, and data analysis is that the understanding and communicating about these has become part of the basic education in psychology. Other forms of modelling do not have this shared understanding, or even a shared vocabulary which to describe either their methodology or implementation. Aside from the difficulties in communicating results, this means that anyone using other modelling approaches must also provide much more detail about how the model is designed and assembled. Less familiar methods invite explicit consideration of the form of modelling used.

While this is valuable no matter what method is used, it may mean that alternative forms of modelling are judged on different grounds to existing methods.

The mathematical nature of many psychological models is disguised by their shorthand expression, reporting isolated, but not meaningless, parameters. The linear equations behind these mathematical models are rarely stated explicitly. In part, the covert nature of mathematics in psychology has developed because the use of a small set of standardised statistical methods used allows writers to assume that readers will fill in the details of the model for themselves. The underlying logic and assumptions are well enough established that writers expect that they will be understood by all academic psychologists, and so do not need restatement.

While it is not unreasonable to assume that informed readers should be able to fill in the gaps, the result is that not only the underlying mathematics, but also the modelling itself, is cloaked behind this codified shorthand presentation. Researchers and readers are able to distance themselves from the mathematical nature of many models, possibly insulating themselves from concerns that they might otherwise harbour about the form these models.

Although quantitative research in psychology relies on a series of modelling steps, much of the modelling carried out in psychology is somewhat covert. A common modelling process can be used to illustrate this. This process has a number of distinct modelling steps, involving various model forms. It begins with a theoretical description, from which a box and arrows model is extracted. This model is either described verbally, or as a diagram showing linkages between constructs. In the second step, the model is translated into a formal mathematical statement of a hypothesised relationship between constructs. Although this form of modelling is routine practice in social psychology, it is not always visible. Parts of the model are often not explicitly stated. For example, the model may be located in a series of statements about hypotheses to be tested, possibly separated by sections of justifying commentary. The corresponding mathematical model is rarely stated at all; usually only the parameters of an equation are reported. Sometimes not even these are reported, only the significance of the test. In effect, this equates to reporting that the data are not inconsistent with some (unstated) mathematical model.

Part of the difficulty in identifying theory and modelling in the social sciences arises from misunderstanding the process in physics. This is coupled with a temptation to idealise this misunderstood process. For example, the formal and elegantly simple theoretical equation $e = mc^2$ is sometimes assumed to underlie the theory of relativity (Shoemaker, Tankard, & Lasorsa, 2004). In fact $e = mc^2$ is the result, not the theoretical foundation. The theoretical idea that led to special relativity can be described by a simple verbal statement: the speed of light and the rules of physics are constant, and hold even when the frame of reference is moving. From this Einstein, developed an analogy based on moving vehicles, and a mathematical model, that reduced to the famous equation.

Computer Simulation

Computer simulation is a particular form of modelling. It is used in many disciplines, for a diverse range of tasks. A familiar form of computer simulation is as a technological tool, such as the simulations used in laboratory demonstrations in psychology. These simulations do not have scope for experimental exploration; they are expected to perform a set of tasks in a predictable and reliable manner.

A different use of simulation is as a research tool. In this form, simulations have potential for exploration and experimentation, because their design allows scope for the model to produce unexpected outcomes. Experiments can be generated from these simulations by varying environmental conditions, or by changing some elements within the model, or by changing the concepts being simulated.

The most unpredictable outcomes generated by computer simulations are particularly characteristic of target phenomena with nonlinear characteristics and, especially, by large systems with such components. Trajectories, and therefore outcomes, in nonlinear systems are very dependent on initial and boundary conditions. The behaviour of simple, solitary, nonlinear components may be intuitively manageable, but as components are added this rapidly ceases to be the case. The result is that outcomes from nonlinear systems, and especially large nonlinear systems, may not be obvious. Beyond this, these non-obvious outcomes can exhibit surprisingly regular emergent patterns. In general, nonlinear systems are not tractable through analytical techniques. Computer simulations are the primary means of exploring nonlinear systems, including large, complex, nonlinear systems. They allow many repetitions, so that initial and boundary conditions can be varied to determine how possible outcomes might be distributed, or initial and boundary conditions can be controlled while the model is experimented with. For example, a simulation model of the transmission of a belief might begin with the initial prevalence of that belief set at 1%, 5%, 10%, and 50% of the population. The belief might initially be evenly spread, or there might initially be clusters holding to the belief. The belief might be fixed at the external boundary to the model.

Realisation of a model in a simulation allows the researcher to go beyond what is already known of a system. Beginning from a base of theoretical knowledge, computer simulation may allows us to extend the scope over which we might apply that knowledge and to explore the implications of theory beyond its immediate reach. In the social psychology setting, this means that computer simulation may allow us to extend our knowledge about individual psychology and groups, so that the implications of these in larger populations might be explored.

As with modelling more generally, simulation in the social sciences faces some particular issues that simulation in other disciplines does not. Simulation is used widely in the physical sciences and engineering, disciplines that have the luxury of deterministic mathematical models of the processes being modelled. For example, given the loads on a structure and a knowledge of the strength of the materials, civil engineers can use simulations to predict the response of structures to earthquakes. It is also used in the biological and ecological sciences, where processes may not be deterministic, but where important parameters occur within a relatively small range of possibilities. As another example, lions have a restricted range of possible reproductive lifespans, reproductive rates, and food consumption. From these known quantities ecologists can predict the stability of a population of lions.

The social sciences are less obviously amenable to simulation, at least through exclusively mathematical models. Humans generate a complicated mix of relevant factors, interactions, confounds and unknowns. Nonetheless, computer simulations have been applied to the social sciences since the early days of computing. Despite an initial period of high expectations, the outcome of some early simulations, notably economic and demographic models, was disappointing. These early models failed to generate successful predictive models (Halpin, 1999). From a position of hindsight, this failure is largely explained by the mathematics of complex nonlinear systems. The sensitivity of the models to initial and boundary conditions largely explains why these models did not produce successful predictions and, indeed, why it is unreasonable to expect that single runs of these models will produce reliable predictions under all circumstances (Haag & Kaupenjohann, 2001).

As with the broader social sciences, computer simulation modelling seems to offer much for psychology, but has not been widely used. One factor that might restrict uptake of simulation is that there is little familiarity with this approach, as computer programming skills are not part of most social psychologists' training.

A second factor might be that computer simulation tends to cross disciplinary boundaries. It has been adopted most widely by the disciplines for which the primary focus is at a more aggregated level, thus suiting the research questions generated in these disciplines. But simulation models require theory about individual-level functioning. This can result in simulations based on individual-level characteristics that are derived from a researcher's guess about individual behaviour, but that would not be supported by evidence from psychology. For example, many economic models are based on an assumption of individuals making decisions in an exclusively rational way. Thus, in addition to the potential for simulation modelling in psychology, there may also be potential for psychology to contribute to modelling in other disciplines.

A third factor might be a "You can't do that!" reaction. This reaction, that there is something inherently wrong with computer models in psychology, possibly has its basis in a confusion between using computers simply as tools or using computers as metaphors (Hastie & Stasser, 2000). Gigerenzer (2000) has noted that we tend to think about psychological phenomena in terms of the tools of research, and our tools can act as a covert constraint to thinking. The digital computer might be one of the better candidates for illustrating this, as it is almost a cliché to use the digital computer as a metaphor for the brain and cognitive function (for an example from popular science writing see (Pinker, 1998)). Further some styles of simulation in psychology do use computers as both metaphor and tool. For example, neural networks are suggested as good models for understanding some cognitive processes.

But the computer as metaphor is also recognised as being limited, not in the least because digital computers are serial, stand-alone, general purpose machines, rather than cyclic, massively parallel, embedded, special purpose machines. Only the simplest brain functions can be modelled directly, and even apparently simple mechanical skills, like walking, have proved very difficult artificial intelligence tasks. The widespread accessibility of computers has revealed these machines as being profoundly stupid, and if a modeller was to claim that the digital computer might provide a good metaphor for social behaviour, a sceptical "You can't do that!" response would not be so unreasonable.

While the computer and some types of simulation may provide a useful metaphor for some aspects of psychology, other types of simulation, such as simulation of the actions of a number of individuals, do not. In these forms of simulation, the computer does not provide a metaphor for social behaviour. It is simply used in its original role, as a device for automating the manipulation of data. Here the computer performs steps that could be carried out with a pencil and paper. The speed of the computer's processing allows the steps to be applied to large arrays of elements, making the realisation of models of groups and populations practicable. Algorithms are written to realise the model, and there is no suggestion that the form of simulation might be helpful in understanding the target system. In this case, the use of computer simulation raises no concerns about the computer's validity as a metaphor in the model, simply because the computer is used only to realise the model and is not itself part of the model.

Another reason for the slow development of simulation as a research method in psychology might be that it does not slot cleanly into any of the most developed forms of scientific methodology (Hales, 1998; Winsberg, 2001). It is neither purely inductive nor deductive, but has elements of each (Hales, 1998). Simulations may not perform well when measured in terms of their predictive capabilities. But while many types of simulation cannot successfully predict what will be, they may be able to predict the range of possible outcomes (Haag & Kaupenjohann, 2001). Although some see this as weaker than prediction, other methodologies cannot identify possibilities.

There are many different types of computer simulation, the choice between them depends on the nature of the phenomenon of interest and our knowledge about it. Beyond minimal commonalities (they all use computer hardware and software) there is very little that could be identified as common among the various computer simulation techniques. They range from numerical solutions of analytically identifiable algebraic equations, to abstract representations of theoretical structures such as neural networks. While numerical solutions are not especially promising in psychology, some of the network models are. Notably, neural networks have become relatively common in models in cognitive psychology and have been applied, more rarely, in other branches of psychology.

The explicit modelling of a network of connected components is categorised as a connectionist model, although some writers reserve this term exclusively for neural networks. In the broader sense, connectionist models include neural networks, cellular automata, and agent-based models. Agent-based models are described in detail in Chapter 4.

Neural networks were originally inspired by the biological model of a network of neurons. While neurons grow and strengthen their interconnection through a biological process, neural networks simulate this process in software. Of connectionist models, neural networks are probably the most familiar in psychology, and especially in cognitive psychology.

A neural network is developed by training it to associate input and output patterns. The network is presented a series of input and output patterns, and the neural network programme adjusts the weighting on the path of interconnections between input and output in such a way that eventually the network reproduces the relationships presented in the training models. Neural networks can predict outcomes, but in an entirely atheoretic way. Training the model is carried out through presenting patterns to the model, and beyond the structure and components of the network itself, no other theory is applied to the model. As the model has no theoretical base to work from, it cannot be expected to respond according to any theoretical principles, or to correctly respond to novel combinations that it has not encountered in training.

Neural networks are at their most useful in demonstrating neural function in unitary processes, particularly those carried out in highly specialised brain structures. For example, neural networks have been successful in demonstrating that neural systems can accomplish pattern recognition tasks. In this mode their research strength is in providing an existence proof. As they are atheoretic, they are less helpful in assisting our understanding of these tasks than in assuring us that our theories and assumptions can work.

Cellular automata and agent-based models are also connectionist models, but they are fundamentally different to neural networks, as they are theoretically based. That is, each cell or agent in the model explicitly behaves in accordance with the theorised process that is happening at an individual level. These individual level representations are collected into a single model of a number of individuals, and so cellular automata and agent-based models produce a bottom-up representation of the aggregation of that process in a number of individuals. Unlike a neural network which adjusts its own connection strengths as part of a learning process, these models are programmed by the researcher.

Cellular automata are variously defined, but a common feature is that at each step, the state of each cell is determined by a decision table based on a history of its own, and its neighbours previous and current states. Cellular automata can be very simple, and in this form they are of interest to a branch of mathematical research that aims to understand how patterns develop in generic cellular automata (Sarkar, 2000). Among the findings of mathematical research into cellular automata is the recognition that cellular automata can be assembled to produce Turing machines. This means that even the simplest members of the family of agent-like machines is a very flexible universal computing devices in its own right.

A second result from the mathematical analysis of cellular automata relates to the possible outcomes from cellular automata. In an extensive exploration of cellular automata, Wolfram (2002) has identified a very small number of possible classes of outcome. While it is not demonstrated that this also holds for agent-based modelling, Wolfram's work provides one possible way to interpret output from agent-based models.

Formal mathematical analysis generally calls for strict definitions. In the case of cellular automata, the definition of cellular automata requires that cells are located in a fixed position, on a map with a regular geometry, and that the states of each cell are determined from a decision table, and that cell states are represented as integers (Sarkar, 2000). But beyond formal mathematical analysis, less formal versions of cellular automata have been used to represent physical, ecological (Gaylord & Nishidate, 1996) and social (Hegselmann, 1998) processes.

Cellular automata have been used for modelling in psychology and the social sciences. Prominent among these applications are models of the spread of patterns of social influence (Nowak, Szamrej, & Latané, 1990) and social support (Hegselmann, 1998). The potential for application of cellular automata to the social sciences is, however, restricted by the incompatible almost defining features of cellular automata vis–vis social systems. Individuals in social networks are not located within neat, homogeneous social networks with Euclidean geometries. Further, the behaviours of the individuals are complex and multidimensioned, and their decision-making often cannot be approximated be simple look-up tables. While this restriction may be tolerable for some applications, notably ecology and population dynamics, it is often too restrictive in a social modelling context.

As the constraints on cellular automaton models are released, they begin to resemble simple agent-based models. The next chapter describes this class of model, and its realisation in software.

Simulation models, statistical models and theory

The way of doing science in psychology has been to take things apart. Having a large collection of isolated parts is not especially interesting, as what we are usually trying to understand is some whole behaviour pattern. There is an implicit assumption that once all of the parts can be isolated, the whole can be reassembled from them. A similar problem exists in cognitive science, where models have been developed for cognition, perception and movement, but without an integrated approach, models that are otherwise not unreasonable cannot be formed into a functioning model. Further, these models do not incorporate the environment, which itself is an essential component in the overall system. This has led some researchers to take an approach in which the overall system is constructed. Constructing a model that is physically located in the environment both allows and requires explicit links to that environment.

Maybe another possible test for a theory is its potential for reintegration into something meaningful. For example, theories of personality purport to provide some description of patterning but attempts to use these as bottom-up components are relatively rare.

Foremost among these issues is deciding on a philosophical foundation to define the roles that multiagent models might take. In the existing literature there are attempts to place multiagent models everywhere from falsificationist to rationalist to constructionist. Axelrod (1997a) has multiagent modelling as a "third way of doing science", with the other two being inductivist (empirical) and deductionist (rationalist). Along almost identical lines we have the Wolfram (2002)claim that cellular automata offer "a new way of doing science".

It seems that whatever agent-based modelling can do, what it cannot do is provide accurate, general purpose, falsifiable prediction. This is only to be expected where we have systems that include any sort of nonlinearity. These are characterised by their sensitivity to initial conditions; a small difference can make a large difference in outcome. We only need one unknown to be slightly different and we will likely seem to falsify the model.This problem is common with other forms of simulation, and for other means of analysing complex systems generally.

In proposing agent-based modelling as a methodology for psychology, we should have some ideas as to the way that agent-based modelling might add to psychological knowledge. Others working with agent-based modelling models, and with social simulation generally, have located simulation in a variety of places philosophically, from empirical to constructivist. This has implications for the verification and validation of agent-based modelling models, particularly where we claim these to be representative of naturally occurring processes.

While some linear processes may exist in psychology, most behaviour is inherently nonlinear, most obviously we may or may not do something, a nonlinear outcome. In combining these individuals, we would expect that the observed of evolving behaviour of a combination of a number of non-linearly behaving individuals will be complex, and sometimes chaotic. The outcomes of such systems is highly dependent on not only the initial conditions, but also on the state of an environment. As a result is is highly unlikely that we would ever have sufficient knowledge to produce a accurately predictive model. These are unlikely to ever produce a model that would satisfy a falsificationist approach to science (Haag & Kaupenjohann, 2001). This is a reasonably active area of publication for many forms of simulation. Similar concerns exist where the simulation uses neural connectionist or nonlinear dynamic modelling. Many of issues for agent-based modelling are common across all types of simulation in a nonlinear dynamic domain.

This area is so new that there are issues attached to almost every facet of agent-based modelling. Possibly the most fundamental for this work is what knowledge we can hope to gain from exploring agent-based models. It is highly unlikely that agent-based models will be testable in through the usual scientific process of testing model predictions against the real world. This is because nonlinear systems are so dependent on initial conditions and inputs from the environment that it would be very difficult to produce a model that was not disconfirmed in many, or even most, trials.

This suggests that simulation modelling may not necessarily fit into a realist science. The act of creating a simulation model is an approach that explicitly constructs in an attempt to understand things that we observe. This is a slightly unnerving situation: computer simulation and the use of mathematical modelling and algorithms are likely to be most comfortable for researchers that routinely use statistical mathematical models, that is for researchers working from a realist philosophy of science. In contrast, computer simulation and modelling seems to fit better a constructivist approach.

On the other hand, simulation offers an alternative approach that can identify the possible outcomes of complex and nonlinear systems (Haag & Kaupenjohann, 2001). Conventional falsificationist science depends on having some sort of theory as a starting point. It has little to suggest how we generate these theories, except as modifications to existing theories that may explain a difference between theory and observed behaviour. Testing these theories is an eliminative rather than a generative process.

Simulation allow us to explore possible alternative outcomes (Haag & Kaupenjohann, 2001). In the real world, the range of these possible outcomes is often outside our experience; our experience is often limited to a single outcome that did eventuate. Similarly, working with simulations allows us to access thinking about whole systems, through letting us observe whole systems. This type of thinking does not necessarily come naturally, and working with computer simulations can provide an opportunity for us to develop some intuition about these processes (Resnick, 1994). This means that simulation has much to offer as a theoretical tool, as it avoids some of the restrictions of our experience and understanding.

Simulation as a research method

As noted earlier, simulation does not fit neatly into mainstream research methods. There are two major elements that differentiate simulation from other research methodologies. One concerns the characteristics of simulation as a research methodology in its own right, whether it is being used to estimate solutions to deterministic equations, such as might be used in the physical sciences, or to model a fuzzier form of theory in social psychology, or to explore the evolutionary development of a faculty in biological psychology.

A second strand concerns the subject matter relevant to simulation in the social sciences, and particularly to agent-based modelling. In the social sciences, simulation is of particular interest for its potential in exploring complex and chaotic systems. In the physical sciences, theory is derived from a small set of fundamental theoretical relationships that can be combined to describe many situations. This can produce very complicated mathematical representations that are analytically unsolvable. Simulation is often used to produce solutions for these situations. The mathematical expressions involved are, in many cases, linear. So the physical systems for which solutions are sought tend to be complicated, but deterministic.

The situation is quite different in the social sciences. Psychological theories have relatively limited scope; they apply in specific circumstances or situations, but small changes in these conditions can produce different outcomes. In natural settings many different psychological processes are active, or available to be activated. A small change in the conditions can trigger activation of different cognitive processes and behaviours, and there is a dynamic interplay between these cognitions and behaviours and the environment. This feature of small changes in conditions leading to large changes in overall outcome is associated with nonlinear systems, and is probably the best known consequence of chaotic, as distinct from complicated, systems. Less well known is that chaotic systems tend to arise out of nonlinear systems (Gleick, 1988) and that psychological systems are inherently nonlinear. Possibly the simplest demonstration of this is that linear systems are, by definition, continuous, while an behaviour might, or might not occur, and there can be an instantaneous change of behaviour state.

Experimental controls in psychology serve to constrain both the available options, and the environment. This constraint allows us to produce an approximation of linear behaviour within a particular range. When released from experimental constraints, cognitions and behaviours are more free to enter ranges where outcomes are no longer so predictable. Unstable states become possible, as do a variety of different stable states. Exploring the range of possibilities that can develop in naturalistic settings becomes a more tricky task than understanding how an individual functions in a controlled setting. To do, so we need to use tools that give us access to complex and chaotic systems.

If complete chaos were the only outcome of a complex system, there is little that we could learn from studying it. All we would know is that the system will collapse into chaos, where it will produce outcomes that, although deterministic, approximate randomness. While the mathematics of this is of interest to complexity science (Morcol, 2001), it is of limited interest in the social sciences.

The social sciences in general are based on observing patterns that develop in natural social systems, and trying to understand these patterns. An interesting feature of complex systems is that they can, paradoxically, produce patterns (Cohen & Stewart, 1995) from very simple rules. These outcomes are not direct reflections of the processes generating them, and so are not obvious consequences of a large number of individuals executing processes. These patterns in outcomes are labelled emergent outcomes. Emergent outcomes are, in some sense, surprising, although surprise suffers obvious problems when used as a defining characteristic of emergent features.

These features of nonlinear systems, sensitivity to initial conditions and surprising outcomes, make predicting the outcomes from nonlinear systems risky, especially in social systems that are very exposed to a broad range of external influences.

The gold standard for a theory has been its predictive value; a good theory should both explain what is known of a phenomenon and be able to predict other, previously unobserved, phenomena. This conception of what makes a good theory is that it is essentially a stand-alone theory. As such it does not require any interface between other good theories functioning in related areas (Axelrod, 1997a).

A feature of simulation is that it has a different relationship with theory from other methods. In simulation, we do not develop theory from an attempt to explain observed real world phenomena. Nor do we develop theory as the logical outcome of a predecessor set of given facts. Simulation is neither purely inductive, nor purely deductive, but may include inductive and deductive steps (Axelrod, 1997a).

Simulation allow us to explore outcomes and in doing so it allows us to develop our intuition about these processes. This means that simulation has much to offer as a theoretical tool, as it avoids the restrictions of our experience and understanding.

Halpin (1998) has proposed a set of possible routes to theory development used in simulation modelling. These are essentially permutations of a sequence of activities: Assumptions (theory), Runs, Observations, and Explanations.

Features of simulation modelling

Choosing simulation as a research method forces a discipline in being specific in representing the construct. This is not exclusive to simulation as this discipline is also necessary in other forms of modelling. Conventional modelling depends on substituting measurable variables for constructs and on defining the relationship of the constructs and, in turn, the variables. Although rarely explicitly stated as such by researchers in psychology, statistical models of these relationships usually take the form of mathematical equations. Thus there are two components to the representation process in statistical analyses: constructs are represented by measurable variables and their interrelationship is represented by a mathematical equation.

Simulation also requires a representation of constructs and their interrelationships. The form of representation of constructs in simulation differs in a fundamental way from their form in empirical research. In empirical models the construct is represented through a variable that can in turn be mapped to a tangible measure. Hypothesised relationships are then directly mapped from the theoretical to the measurable.

In simulation models the representation of the construct remains theoretical. The simulated representation of the theoretical construct must be to be realised in an algorithmic form and so requires a detailed statement of the construct.

Most models in psychology represent interrelationships in the following form: if there are A more units of variable x we expect to find B more units of variable y. These statements have nothing to say about how a change in x and a change in y are linked, only that they are. At best these models specify the direction of causality.

The representation of interrelationships in agent-based models is much more detailed than in statistical models. For example, statistical models of trust might assemble a set of variables that theory suggests may be correlated with trust, and test whether some function might link these with trust at a population level. An agent-based simulation model of trust requires theory that specifies how each individual comes to a point of trusting. This involves theorising what information the individual can obtain, how they might assemble that information, and how they might come to a decision. Further, an agent-based model might need to specify the credible variation between individuals for each process. All simulation models require a precise definition of the processes linking constructs. The elaboration of these processes is likely to be more detailed than the constructs themselves. Further to this, agent-based models using multiple agents also require specific statements of the linkages between the individual agents.

At a conceptual level, the model-building process is common for research using conventional methods and for simulation. Theoretical constructs and their relationships are converted to a form where they can be manipulated and measured. The approaches diverge in the form of detail resulting from that conversion. At the realisation stage of their development simulation models are quite unlike statistical models, both in form and in the balance of constructs and processes represented.

The models arising from simulation and statistical modelling differ in their relationship with theory. Simulation models retain a close relationship with theory, and can contribute to the development of the theory. Statistical models disconnect themselves from theory, and their results merely allow us either to reject or not reject the theory. Beyond this they do not contribute to further development of the theory.

Chapter 4

Agent-based Models

Agent-based modelling is a computer simulation technique that is characterised by agents. In agent-based modelling, agents are small computer programmes that represent discrete elements in a simulation model. Each agent has a set of characteristics, and a set of actions that it can carry out. The most obvious use for agents in a social psychological setting, is for representing individual people.

The technique of agent-based modelling promises almost limitless possibilities for a researcher, as agent-based model simulations can incorporate anything that can be programmed. Many techniques are readily available, prepackaged into agent-based modelling packages, as these provide access to a variety of artificial intelligence devices for knowledge representation, manipulation, learning, and decision-making. In a research setting, this translates into a very flexible facility for representing psychological concepts and processes, allowing us to represent situations that are otherwise impossible to explore through more traditional means, through explicit simulation of agent interactions and sequences of actions.

This chapter introduces software agents, and agent-based models. The chapter begins by defining agents and agent-based modelling. Agents can be assembled and used in a number of ways, and the chapter continues by describing the three major ways that software agents can be assembled and used: as single agents; as agent-based models; and as engineered multiagent systems. These are described, and the features of multiagent systems are broadly outlined. The features of agent-based models are discussed in more detail.

Having defined what agents and agent-based models are, the chapter goes on to describe how these models are constructed, and the uses to which they might be put. This section also identifies and reviews the major software packages that are available for constructing agent-based models, and the key features that are offered in these packages.

Agents

Software agents are computer programmes, or sections of a computer programme, that act in pursuit of their own goals (Wooldridge, 2002; Jennings, Sycara, & Woolridge, 1998), independently of outside direction. This ability to act independently suggests that an agent should take action without being told what to do. To do this, they need to collect information, and to determine a course of action based on this information. Agents are located within an environment, often a larger body of programme, from which they can gather information. This environment may also accommodate other agents. An example of an agent might be given by airfare search agents. These agents are located in a virtual environment with access to the Internet, from which they obtain information on airfare pricing. The information found is returned to a segment of code that can make comparisons and an eventual decision on the best offer.

The word often used in the agent literature to describe this independent information collection and action is autonomy (Jennings et al., 1998; Wooldridge, 2002). In truth, full autonomy is challenging to implement in artificial intelligence and so, in practice, there is a wide interpretation as to the degree of autonomy required for a segment of code to qualify as an agent. An agent may be as simple as an algorithm that collects information and applies a programmed action in response to information that is available from the environment. Earlier definitions of agents also required that individual agents be flexible (Jennings et al., 1998), a requirement that demands that each agent has a degree of inbuilt adaptiveness. More recent definitions, such as that given by Wooldridge (2002, p. 15) does not impose a requirement of flexibility on individual agents. Rather, he suggests that agents are characterised by an ability to perceive conditions in their environment, and to respond to environmental conditions by taking appropriate action to pursue their goals, including interacting with other agents (2002, p. 23).

Within computer science, agent systems are themselves a branch of artificial intelligence. Artificial intelligence is difficult to define. For example, it has been described as "the branch of computer science that is concerned with the automation of intelligent behavior" (Luger, 2002, p. 1). This definition does not differentiate psychological and computer science concepts of intelligence. Rather, it implies that artificial intelligence should be able to emulate naturally occurring intelligent behaviour, leaving open the question as to what *intelligent behaviour* is.

The definition of naturally occurring intelligent behaviour is in the domain of psychology than in the domain of computer science. While not offered as an explicit definition of intelligence, Neisser and colleagues (1996) point to individual differences in ability to "understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77). While this is only one of a number of ideas about intelligence that have been generated within the discipline of psychology, it is possibly one of the more useful for computer scientists hoping to replicate natural abilities and behaviours.

In contrast to these broader definitions of intelligence, classical artificial intelligence has tended to focus on a particular set of abilities. These include knowledge representation, learning, searching for information, problem solving and decision-making. For agents, taking appropriate action may mean that they need to be able to solve complex problems, or make complex decisions. To do so, agents can be complex, possibly incorporating artificial intelligence devices. Complex single agent devices may have some of these abilities built in, but the computational overhead that they require may not be practicable in systems with a number of agents.

The incorporation of individual artificial intelligence features is not, however, a necessary feature of agents. Agents may also be quite simple. Very simple agents, that have not been provided with artificial intelligence features, can be assembled to generate a system that might exhibit some intelligence. Given the right conditions (Bonabeau, Dorigo, & Theraulaz, 1999, p.9), the multiple interactions of simple agents can produce an entity that can perform effective searches, and that can self-organise to respond to environmental conditions. Insect colonies, such as those constructed by ants, bees and termites, provide natural examples of systems in which simple agents can effectively search for food, and construct complex structures. Attempts to construct systems of this type described this as *swarm intelligence* (Bonabeau et al., 1999, p.7).

Single agents and systems of agents

As the previous section has alluded to, software agents are defined and located as a branch of artificial intelligence. Agents are not necessarily located in systems of a number of agents. Some are designed to work alone. A familiar single agent is the Office Assistant in Microsoft Word. The Office Assistant monitors a user's interaction with the Word programme, with the purpose of offering appropriate help information.

Two other types of agent system, agent-based models and multiagent systems, are assembled from a number of agents. Agent-based models and multiagent systems are related, in as much as they are constructed from a number of agents. These systems are similar in many ways, for example, the definition of *agent* given in the previous section is drawn from the multiagent systems literature. There is also traffic in ideas between the users of agent-based models and the designers of multiagent systems: engineers constructing multiagent systems have used explorations of naturally occurring social organisation to inspire efficient algorithms (Ray & Liew, 2003), while the software technology that is used to develop agent-based modelling packages is rooted firmly in multiagent systems, and draws on developments in intelligent systems.

The difference between agent-based models and multiagent systems is rooted in their purpose. Multiagent systems are designed to distribute simple tasks among a number of agents, in order that the whole system can carry out a more complex task. That is, multiagent systems are designed to carry out some function. This functionality can only be assured if the systems are reliable, which in turn depends on them behaving predictably in the range of environmental conditions that they might encounter. Agent-based modelling systems are, as the name suggests, used for modelling and experimentation through simulation of the individual elements in a system. These differences in purpose between multiagent systems and agent-based models in turn result in differences in how each is designed.

Multiagent systems

While the design of multiagent systems may have been inspired by natural systems, there is no need for them to be representative of any natural system. The function of individual agents may theoretically even be completely abstract, as long as they work together to carry out the overall system's designated task. Engineered agent systems often use artificial intelligence devices. These artificial intelligence devices may have been developed as attempts to replicate naturally occurring intelligences, but their development has been in pursuit of functional devices, rather than an intent to be representative, or to mimic natural phenomena.

As systems that have been engineered to carry out tasks, multiagent systems need to be reliable, and their behaviour well understood. This is often achieved by limiting these systems to small numbers of simple agents, restricting communication, limiting cooperation, and implementing centralised control (Edmonds, 1998). Designers of multiagent systems avoid their systems exhibiting the very phenomena that users of agent-based models seek - novel and emergent properties (Edmonds, 1998).

Agent-based models

In contrast to multiagent systems, agent-based models are intended to be representative, and their very purpose is to explore complex system behaviours. Agents are designed to explicitly represent the individual elements in a naturally occurring system. In a psychological setting those individual elements might be the individual people in a population. The degree of sophistication of the individual agents in such a model should depend on the representation demands of the model. Agent simplicity is desirable, both for producing a parsimonious research design and for the practicality of maintaining a computable model with a large number of agents. Balancing this, agents need to be sufficiently complex to represent essential features (Edmonds & Moss, 2005), without being too complex to understand, debug or run. The segments of code that form individual agents may be relatively simple, not much more complex than cellular automata. Even models as simple as cellular automata can produce informative results, for example, agents that are only a little more complex than cellular automata have been used to model the development of support networks among mobile populations of individuals with differing risk profiles (Hegselmann & Flache, 1998), the spread of attitudes through a network due to the influence of neighbours (Latané & Nowak, 1994), and the spread of cooperation against predators in the presence of freeloaders (Jaffe & Cipriani, 2007).

In an agent-based model, a number of agents are assembled into a model that allows individually functioning agents to interact with each other. Structurally, this results in a high degree of interconnection of a number of elements. High degrees of interconnection are characteristic of three types of modelling device: neural networks; agent-based models; and cellular automata. These interconnected models differ in the types of systems that they are best suited to modelling. Neural networks are best suited to low level processes, for example demonstrating that interconnections between a large number of very simple neurons can carry out complex pattern recognition and learning tasks. Cellular automata are suited to exploring systems in which individuals make simple logical decisions, influenced only by their own state and by the state of close neighbours, for example an individual may move closer toward others if the population density is low and there are vacant neighbouring locations. The definitions of both neural networks and cellular automata have rules that force simplicity in their elements and that restrict patterns of interaction with neighbouring elements. Agent-based models do not have these constraints, they consist simply a number of self-directed agents that act autonomously and are interconnected. There are no constraints on the pattern of any interconnections in agent-based models.

The usefulness of agent-based models arises in the overall behaviour of the system being, to some extent unpredictable. An agent-based model typically consists a large number, possibly hundreds, of relatively simple agents. The large number of agents and their interconnectedness tend to combine to generate systems with complex behaviours. As discussed in Chapter 2, these large, interconnected systems of nonlinear elements can behave unpredictably, and are not amenable to analytic solutions. Further, they are so sensitive to initial conditions, and to conditions at the interfaces with the wider environment, that small differences in these can produce completely different outcomes. The combination of these means that we should expect that social systems with a number of individuals are not necessarily amenable to standard analyses. One form of analysis that can access these systems is simulation. The availability of powerful computing facilities has made simulation of complex systems possible. Agent-based model are one form of simulation that can be used to access systems of interacting individuals.

Agent technologies and artificial intelligence offer a tantalising degree of sophistication: a number of smart features can be added to models. Many agent-based modelling packages include libraries for knowledge representation, problem solving, and learning. Theoretically, agent-based models could also have rich inter-agent communication, but this is a much more challenging technology that is still being developed. The availability of these features means that we can represent many features of human cognitive function.

In a complete model, the agents occupy a virtual environment that consists of the agents themselves as well as a section of programme that forms their environment. Within the model, agents and their environments are realised through separate segments of code. For example, in a model of traffic in a large city (Bonabeau, 2002), the environment represents the transport infrastructure: roads and public transport. This environment is occupied by individuals who carry out a range of daily activities that require movement within the city. In the model, a population of simulated citizen agents go about their daily activities in the simulated city environment, with their movements tracked minute by minute.

The environment may be also be non-physical. For example, in a model of on-line trading, the environment might represent the reputation information and transaction management that is made available by on-line trading sites, while individual traders are represented by agents.

In addition to modelling the environment, main body of an agent-based model programme manages the social environment through managing the interaction of agents, such as allocating turns for the agents to act. An example of an agent-based model in which the environment consists only of other agents is given by Mosler (2001; 2006), who has constructed a model based on the Elaboration Likelihood Model of influence in a small group. In this simulation, the environment only contains the five interacting agents (Mosler, 2006).

Agent-based modelling is thus based on simulations of the autonomous behaviour of a number of individual agents, interacting within a simulated environment. These agents interact with each other; their actions depend, in part, on the actions of the other agents that they interact and exchange information with. Therefore, these models are inherently social, to the extent that agent behaviours are located within a context of the actions of other agents. The explicit social nature of the model may make these models a promising candidate for modelling in the social sciences.

Agent individuality

Within an agent-based social simulation, the agents may be identical, they may have individualised characteristics, or they may be qualitatively different. Less obviously, agents might also represent larger entities, such as organisations. Other agents might represent largely automated systems, such as the auction management system on an internet trading site.

Identical agents have been used to generate artificial life models. Reynolds' (1987) model provides an example, modelling of the flocking behaviour of a group of animals. The agents in this model, which Reynolds dubbed *boids*, were equipped with three simple steering rules: avoid crowding other agents; move toward the centre of the flock; and move toward the average direction of the other agents. The use of identical agents in this model demonstrated that flocks of birds do not need a bird in the role of leader, rather flocking behaviour arises within a group of identical agents that shares these rules. Similar models of the movement of identical agents have been applied to building evacuations, showing that the layout of a building can lead to bottlenecks in an escape path as people's movement becomes disorganised around constrictions in the building layout (Helbing, Farkas, & Vicsek, 2000).

More often, agent-based models use agents that have an identical structure, but where there is individual variation within that structure. Examples are provided by the previously mentioned models of social influence (Nowak et al., 1990), in which agents carry individual attributes for persuasiveness and supportiveness, as well as the attitude status of interest, and of the development of social support networks (Hegselmann & Flache, 1998), in which agents carry different levels of vulnerability.

Agent-based models also have the flexibility to represent qualitatively different entities. This is seen in a model of a marketplace that includes customers, suppliers, and reporters who gather and disseminate reputation information (Hahn, Fley, Florian, Spresny, & Fischer, 2007), which is an example of a model with qualitatively different agents. In the course of this simulation, customers attempted to assess a suitable supplier, using a number of criteria. Following the award of a contract, suppliers who did not have enough capacity to carry out the contract, could subcontract through negotiating with others to meet their obligations, or they could default, and possibly spread false reputational information. The reporters interviewed customers and suppliers, collecting reputation information, which they could sell to other agents. Each of these types of agent was allocated different aims, and was equipped with different abilities. In this case, the agents, might represent corporations, individuals, or a mix of these.

If the behaviour of a group of agents is well-enough understood, the group might, in turn, be represented by a single agent. In some cases, a formal description of the behaviour of the aggregated entity might be derived from a formal description of the individual agent's activities (Bosse & Treur, 2006). Less explicitly, Hales (1998) has reported a preliminary investigation in which a simulation was run that developed an identifiable cultural grouping within a population of agents. The agents' behaviour was observed as they developed ways of dealing with these different groupings.

As noted earlier in this chapter, each agent in an agent-based model can be unique: agents can be set up with different initial conditions and different individual characteristics. These characteristics can change during a run, which means that an agent can develop during a simulation allowing individual agents to adapt and learn. Agents can interact with any aspect of their environment, including any other agents. This interaction of individuals is characteristic of social behaviour, and we might regard agents as devices that might be used to model individual social beings.

Artificial societies

A major purpose of agent-based modelling is to generate models of societies. This is exemplified by researchers working with agent-based modelling, who note that agent-based models might be described as "artificial societies of autonomous agents" (Conte, Gilbert, & Sichman, 1998). Natural societies are complex dynamic systems of a large number of interacting nonlinear elements; a combination has the characteristic that quite general outcomes can result from small differences in initial and boundary conditions. This means that the same assembly can produce a range of possible outcomes, some of which might be qualitatively quite different. Artificial societies are not intended to represent existing societies (Gilbert & Troitzsch, 1999; Hales, 1998). Rather they provide the means to explore the range of outcomes that might possibly arise (Gilbert & Troitzsch, 1999; Haag & Kaupenjohann, 2001) through the interactions of a set of agents with particular characteristics.

Where some of the possible outcomes are costly, it is useful to know that these possibilities exist. It would be even more desirable to understand the system dynamics sufficiently to be able to avoid high cost outcomes. For example, agent-based models may be useful for understanding the conditions under which insurgency might be successful (Doran, 2005). In other cases, a unique or surprising outcome may be possible. For example, networks of alliance and obligation that arise from marriages and birth order seniority in the context of Tongan social rank can act to generate a stable social system founded on unstable shifting relationships (Small, 1999). Such possible outcomes cannot be accessed through other means, for example using statistical models.

There are three angles that we might take in investigating artificial societies. Firstly, much agent-based modelling seeks large scale patterns arising from the actions and interactions of the individual agents within artificial societies. There are a growing number of studies investigating groups, norms, cooperation, and cultures.

Secondly, artificial societies may also be considered as an aggregated entity. For example, using a swarm metaphor, a beehive might provide an example of a unit that is self-organising, responsive to its environment, and adaptable. Agent-based modelling can be used to understand how these entities work.

A third possibility is that an agent-based model might be employed to investigate the development of individual agents. This development may arise when agents learn from their interactions with others, resulting in changes in the agents themselves. This approach has been used to identify optimal strategies for formal games. Optimal strategies can sometimes be found through formal mathematics, but they can also be found through trial and error, or learned. Agent-based models can demonstrate how agents develop optimal strategies. An example of this approach is the evolution of a cooperative strategy, identified as "tit-for-tat", in an iterated Prisoner's Dilemma game (Axelrod, 1997b).

Constructing an agent-based model

While the above describes what agents are, it says nothing about how we might construct such a model. The process in some ways parallels other research methodologies (see Table 4.1). The end result should add to theory, forming the foundation for new experiments and models. In simulation modelling the overhead of these new experiments and models is reduced; the major component of the effort lies in writing the software for the initial development of the model, after which parameters might be adjusted, or the model modified a little, and the process run again.

One of the major ways suggested to validate agent-based models is to de-

Table 4.1: *Parallels Between Conventional and Simulation Research*

Experiment or survey	Simulation modelling
Research question	Research question
Design experiment or survey based on theory	Design simulation model based on theory
Carry out experiment or survey	Run simulation model
Collect data	Collect data
Interpret results	Interpret results

velop the models based on sound theoretical bases (Doran, 2006; Moss & Edmonds, 2005). This is a challenging task (Doran, 2006), but one for which social psychology should be well-placed to contribute. Even in social psychology, where there is a lot of theory regarding individual behaviours, these are not necessarily in the form that can be readily incorporated directly into agent-based models. While psychological theories are usually stated in verbal terms, it is possible to formulate theories in social psychology in a mathematical form. For example, the theory of social impact has been stated as $i = sN^t$ where i is the expected social impact, s is the situation dependent scaling, N is the number of sources, and t is a value that takes account of the decreasing influence of additional sources (Nowak et al., 1990). While this returns a mathematical form of the relationship between resulting influence and the number of sources, the resulting equation still contains elements (s and t) that will require the modeller to make assumptions, approximations or estimates to produce a computer algorithm. In the above example, t produces the effect of decreasing the effect of additional influences, thus t must be less than 1, and is often found to be in the vicinity of 0.5 (Nowak et al., 1990).

The nature of the theoretical content should guide the nature of cognitive processes used by the agents in the model. As noted above, agent-based modellers have access to a range of artificial intelligence devices that might be applied to agent architectures. Doran (2006) lists these as:

> 'sets of variables with associated condition-action rules (incl. "fuzzy" rules)
> artificial neural networks
> behaviours and subsumption
> predictive planner and internal model
> logic systems e.g. BDI [belief desire intention]

hybrid and/or multilayer' (Doran, 2006)

The choice of these should take into account the type of process being modelled. For example, artificial neural networks, or subsumption architectures[1] might reasonably be associated with low level cognitive processes, while a predictive planner might be associated with a high level cognitive process. Where both processes are thought to be important, low- and high-level devices might be combined in a hybrid or multilayer cognitive system.

Agent-based model construction tools

While agent based models can be programmed from scratch, this is difficult. It is also an inefficient use of the researcher's time. For example, functions for setting up agents, for setting up and initialising models, and recording data are likely to be required in any agent-based model. These tasks are standard enough that they can be managed by pre-built libraries of routines. Combined with some way of describing the agents, agent-based modelling packages provide the libraries for the house-keeping routines associated with running the models. Typically, these include the management of information, controlling the agents' turn taking, and the recording and display of information generated.

There are a large number of agent packages available. Many are suitable as general tools for agent-based modelling, including modelling for social psychology. Others are designed for more specialised tasks, or with features that are specific to a particular discipline. An example is SimBioSys (McFadzean, 1994), which describes elements in biological terms and includes explicit phenotype and genotype classes.

There is a range of model details that can be implemented in the various packages. Some enforce simplicity in agent design, while at the other end of the scale there are packages that have facilities for highly sophisticated agent design. This equates to a range from systems that are only slightly more flexible than cellular automata, to those that allow for very sophisticated agents, of which there may only be a small number before the computational task becomes impracticable.

The simplest, and the first, of the computer based models automated John Conway's model, called Life. In Life, the rules define the population density

[1]A form of artificial intelligence in which concrete lower level tasks are collected to carry out higher level, more abstract tasks.

that is needed for a cell to live. Cells are located on a two dimensional grid. A living cell needs two or three live neighbours to live, otherwise it dies. New cells are born in empty cells with exactly three live neighbours. Despite the very simple rules, this model produces complex patterns moving through the whole population of cells (Luger, 2002). These simple models are also applicable to some phenomena in psychology, for example as models of the development of social support networks (Hegselmann, 1998), or the effects of influence on attitude distribution (Latané & Nowak, 1994).

Following on from Life, many packages were developed to model the biology, survival, and reproduction of living things in a population. This form of modelling falls under a field known as artificial life, although the field also includes more explicitly biological models (Luger, 2002). Models within an artificial life framework tend to have the agents located within a geographical environment, represented by various forms of grid, which may also contain resources. Some programmes restrict agents to interacting only with their immediate neighbours on the grid. That is, models using these packages tends to be located in a Cartesian two-dimensional grid layout. This may be appropriate for social systems models that require geographical settings; traffic behaviour or evacuation models are good examples. It is less appropriate for models of social networks to be restricted in this way, as they have non-Cartesian network structures: individuals in social networks may have relationships with individuals that are quite remote. In most programmes, even those that require that agents have a location on a grid, these social networks can still be modelled. Geographical locations may be may be included to satisfy the requirements of the programme, but with no particular meaning within the model.

More recently, a number of agent-based modelling programmes, suitable for modelling in the social sciences have become available. The first was Swarm (Minar, Burkhart, Langton, & Askenazi, 1996). Later programmes include Ascape (Parker, 2000), Repast (North, Collier, & Vos, 2006), NetLogo (Wilensky, 1999), and SDML (Moss, Gaylard, Wallis, & Edmonds, 1998), with Mason (Luke, Cioffi-Revilla, Panait, & Sullivan, 2004) a newer entrant.

An early package: Swarm

Swarm was one of the first agent-based modelling packages for general use. It was developed at the Santa Fe Institute, where computer models were being used in a broad programme of research into complex dynamic systems. There was a concern that computer programmes were written anew to simulate each new research idea. This meant that computer models developed in research were difficult to reproduce (Minar et al., 1996). Further, results generated from the programmes might be artifacts of the software design, rather than reflections of the system being modelled (Minar et al., 1996). This is more likely to happen where non-specialist software developers are writing the programme, and where the programme is not subjected to rigorous testing and verification. This situation would apply to most researchers in the social sciences.

Agent-based modelling packages offer a number of advantages. Researchers can increase software reliability, because much of the housekeeping work of agent management, input, output and data display, at the same time making it easier for researchers to develop their models (Tobias & Hofmann, 2004).

With a number of packages available, researchers need some criteria for selecting an appropriate package to use. Criteria might include ease of programming the model, flexibility, the availability of existing libraries, traceability, and whether the package has an active development and maintenance community.

Swarm models are written in Objective-C, which may be a more natural fit with agent-based modelling than Java (Railsback, Lytinen, & Jackson, 2006). That said, Objective-C is not a widely used programming language, although a newer Java overlay has recently been added. Swarm is open source, which means that the functioning of programme itself is traceable. Swarm has been passed to an active development community for continued development and maintenance.

The Java-based packages: Ascape, Repast and Mason

Ascape, Repast, and Mason are packages for developing agent-based models in Java. Although Java is a relatively complex language, it is widely used, so there are many resource libraries available for Java. Tutorials and support are easily found - an important consideration for researchers who will carry out their own agent programming. Ascape was developed at the Brookings Institute, but is no longer under active development. It is not an open source package, and the software has not been passed to a user community, and so will have no further development and maintenance. Repast was developed at the University of Chicago. It is open source, and has been passed to an active development community.

Of the Java-based frameworks, Ascape requires that agents be located within a grid or lattice. The grid or lattice is, itself, an agent, and contains the rules for interactions between the agents. While this is suitable for some simulations, it is too restrictive to represent the topologies of many social networks, other than those in which homogeneous agents are restricted to interacting with their physical neighbours. Despite Gilbert and Bankes' (2002) comment to the contrary, Repast has no such constraint, and is more generally applicable to more general social system structures.

NetLogo

NetLogo models are written in a variant of the high-level language, Logo. Like Ascape, NetLogo locates agents geographically, but NetLogo is flexible enough to readily allow models of other forms of social network. NetLogo has many high-level data management and display functions built in. NetLogo has the unique feature of being able to run a simulation involving a number of different computers linked in a network. Such a simulation allows a number of people to attempt to cooperatively achieve some task by operating their own section of it. This type of modelling, using another Logo based platform (StarLogo), was reported by Resnick (1994). In Resnick's simulation, a number of participants wrote segments to control traffic lights, that were networked to create a city with a number of traffic lights. The overall purpose was for the group to control traffic in a city, with each participant managing their own sector.

This multiprocessor approach also has the important feature that the agents can function in parallel. Where a programme is run on a single processor, agents must function sequentially, rather than in parallel. Most agent-based modelling programmes use an approach in which the agents take turns to act and interact. It is possible that this turn-taking, a requirement of the modelling package itself, rather than the model, affects the dynamics of a simulation model.

Summary: Selection of the Repast package

When looking for an agent-based modelling package with which to carry out the modelling work for this thesis, I found that a number of packages were currently available. I wanted a package that was relatively easily learned and used, actively supported, and in which the behaviour of the package was traceable. Further, I did not want the design of the model to be compromised by the requirements of the modelling package. In particular, it should allow agent interactions with agents who were not near neighbours.

With most packages, agent-based modellers will have to learn both a programming language, and the features of the package itself. Like most psychology researchers, I am not a professional programmer, although I was sufficiently confident with the use of other programming languages to be reasonably confident about being able to develop sufficient skills to develop models in one of these languages. I would have to learn one of three programming languages: Logo, Objective C, or Java. Logo, used in building NetLogo models, is reputed to be an easy programming language to learn, although I have not found it particularly easy myself. Logo is not widely used outside primary school teaching. Similarly, Objective C is not widely used. This has disadvantages, both in finding other code libraries to extend models written in either of these languages, and in obtaining support for learning the language.

Java is a much more widely used language, and as a result there are many resources available, both for learning and for reusable source code. For instance, Java libraries are available for agent communication, learning, knowledge representation, and decision-making, any of which might be useful in agent-based modelling. There is an example of this in the Trading Model (Appendix 1). One section of code, the *Stream* class, assembles data for sending to a file, has been based on code from a Java textbook.

Of the agent-based modelling packages using Java for programming the model, Repast had the most features (Tobias & Hofmann, 2004), and offered the most flexibility at the time that I was starting to develop the model. At the time the Mason package was not yet available. Repast also offers versions through which models can be developed in the simpler higher-level languages, .NET, and Python. Repast has an active development and user community, as do Swarm and NetLogo, which means that the future of the language is assured. It is also open source, this means that the programme listings are distributed with the agent-based modelling package. This has the important advantage for research that the behaviour of the programme is traceable. Last, but not least, it is available for a number of operating systems, and it is free (as in beer). The trust models that I developed for this thesis were developed using the Repast agent-based modelling package.

Reasons to use agent-based models

Agent-based models, and simulation models in general can be put to a variety of uses. As with conventional research approaches, they can be used to gain understanding. In the case of agent-based models, they are suited to developing an understanding of complex processes (Gilbert & Troitzsch, 1999), that involve a large number of people interacting, and having an effect on each other. These processes are difficult, if not impossible, to observe explicitly. Further, the size of social structures can be such that it is difficult to detect what is happening, much less why it is happening.

Agent-based models can include agents with individual characteristics, so diversity can also be modelled directly (Lansing, 2002). This gives a means both to understand processes at an individual level and the effects of diversity more generally. For instance, it can be difficult to incorporate all the dimensions of diversity in a single piece of research using statistical models before the dimensions swamp the degrees of freedom available.

Agent-based models can be used to predict outcomes, but not in a deterministic way (Gilbert & Troitzsch, 1999). That is, agent-based models can predict a range of possible outcomes, but cannot make a firm prediction about which of these will occur in a particular situation (Haag & Kaupenjohann, 2001; Lansing, 2002). This is because the models are large systems of interconnected elements, which quite possibly have nonlinear characteristics. The reasons for this will be given in Chapter 2.

The process of generating the agent's description and abilities involves an explicit theoretical formulation (Gilbert & Troitzsch, 1999). Much of the theory in social psychology is contained in a verbal form, using fuzzy terminology. Programming an agent demands converting a verbal and possibly fuzzy theory into an algorithmic form. Effectively, this is a forced formalisation of theory. As a result, this formalisation process may expose a need to make assumptions about the specifics of theory, which can lead to ideas for more conventional research. This process involves working with theory in an in depth way. This work with theory continues through the process of working with the model.

Conclusion

Agent-based models are devices for computer simulation, that allow the simulation of individual elements in a social system. Within a model, agents are explicit expressions of theoretical ideas about how the individual elements function. That is, the representations of individual agents are demanding of theory about individuals in social settings. This individual level social theory is the natural domain of social psychology, and social psychology is the natural source for the theory essential for building valid agent-based models. Applying theory to individuals in this way, and modelling the resulting interactions of the individual agents has implications that will be further considered in Chapter 5.

Chapter 5

Agents in psychology

The previous chapter introduced agent-based modelling as one method by which the large complex nonlinear systems that are social systems might be researched. This chapter goes on to consider what agents and agent-based modelling might mean in a social psychological context. The most obvious use of agents here is to use them to represent individuals.

The use of agents to represent individuals is not their only possible use. For example, agents have been used to model the individual temples that form a decentralised system for managing irrigation water in Balinese rice farming (Lansing & Miller, 2003). They have also been used to model more abstract constructs, for example the relationships between people rather than the individuals. In my Masters thesis, I used agents to model the effects of increasing stresses on relationships that might arise as a results of increasing income differences. While not explicitly an agent-based model, the elements modelled were essentially agents. The agents in this model represented the characteristics of the relationships between people, rather than the individual people. This approach is similar to approaches using equivalent formulations in engineering, for example choosing either a node current or loop voltage formulation for analysing electrical circuits. Outcomes, might be at an individual level, or at the level of groups, organisations or the whole population.

Nevertheless, while single agents may represent other things, in the social sciences they are most likely to be used to represent individual people. In agent-based modelling, these people are located within a network of interacting people. The entirely general characteristics of nonlinear systems suggests that the behaviour of a system of individuals often will not be a simple summation, either of the behaviours of a number of individuals, or of a number of different behaviours that might be exhibited by any one individual.

This effect of the aggregation of the activities of a number of individuals has implications for the builders of multiagent systems, who need their agents to perform a variety of social tasks: to compete; to negotiate; to communicate; and to cooperate (Wooldridge, 2002). Those engineering such systems need to understand how social networks form and operate, and may need to understand the system behaviours that may emerge from the interactions among agents. The aggregated behaviour of networks of interacting agents also has implications for cognitive scientists. Sun (2001) notes that as well as builders of multiagent systems needing to understand the social science functioning of their devices, cognitive scientists need to understand the cognitive processes of individuals who are socially located.

Social networks are also important in psychology, and techniques exist for exploring phenomena in social settings, or associated with numbers of individuals in groups or larger organisations. These are, however, usually located within each of these levels investigating individuals within social settings, within groups, and within organisations as single entities, rather than as systems. Agent-based systems may provide test beds for exploring the phenomena across these levels, from individual perceptions, motivations, actions and interactions in social settings, and the collective social environment formed by a number of individuals.

This chapter considers the relationship between agent-based models and psychology. There are two main aspects to consider here. Firstly, psychology might provide a source of sound theory to be used in the design of the agents within an agent-based model. Secondly, agent-based models might be used as a device for investigation in social psychology. Although this second is the primary focus for this thesis, it is listed second here because models directed towards enquiry in social psychology are as dependent on a sound basis for agent construction as is a model in, say, economics. The two are not independent.

The chapter begins by considering some issues in the use of individuals as the basis for a model. Both psychology and sociology have points of concern in delineating the relationship between individuals and their environment. These have a slightly different flavour in sociological deliberation to that in psychology. In sociology the concern relates to the appropriate unit of analysis; should structural factors be defined, explained and reduced to processes at the the level of individual, or should structural elements be considered as entities with existence and explanatory powers in their own right?

Designing an agent requires thinking about what might lead an agent to act, and about what course of action the agent will take. The chapter goes on to consider the relationship between the understandings of how individual people behave as agents from the perspectives of psychology, philosophy and computer science. There are some commonalities, in part because the theories are all sourced from folk psychological theories of how people generate an intention to act.

Agent-based modelling provides the opportunity to represent individual characteristics. Where these are factors that might influence the actions of the agent, they might be thought of as personality factors. Psychology has been concerned with identifying and measuring personality characteristics. The chapter considers the potential for incorporating these, as well as social cognitive perspectives on personality, into models.

Agents, social structures, and the transition between levels

Agent-based models have been used increasingly widely in sociology, anthropology and economics. Agent-based models are, by nature, bottom-up models, almost invariably based on modelling individual people. Where a model is built using agents that represent individuals, each agent encapsulates theoretical ideas about how individuals function. Whether the overall model is psychological, sociological, anthropological or economic, the model is built on ideas about the processes at work in individual agents - their perceptions, motivations, cognitions, and actions. Individual processes are, in general, the legitimate domain of psychology.

Agent-based models are usually specified and developed at an individual level, while outcomes may be at any of a number of different levels, including emergent patterns that may be identified as structures. One of the things that differentiates the disciplines of psychology and sociology is the level at which analysis is carried out. The combination of both individual components and structural outcomes in one model means that an agent-based model may be neither entirely psychological nor entirely sociological. Agent-based models have the potential to cross disciplinary boundaries, and this also means that an agent-based model may therefore encounter theoretical challenges from within both disciplines.

Psychology is largely an individualistic science, often investigating people in isolation. While this is less often the case in social psychology, which by definition is concerned with the psychology of individuals in social settings, this too tends toward the individualistic. On the other hand, sociology is concerned with society and social interaction, and inhabits a domain that ranges from the individual to large social structures. There is a fundamental debate in sociology between those who maintain that structural features are generated by individuals, and would not exist without individuals, and those who maintain that while social structures do not exist independently of individuals, they are not necessarily reducible to an individual level.

With its tradition of interest in individual behaviour, psychology is a potential source for supportable fundamental theoretical bases to be applied to agent-based models. Despite this, agents are rarely explicitly designed in terms of explicitly psychological theory. There are a number of possible reasons for this.

Psychological knowledge does not necessarily take a form that allows it to be incorporated into a model of an individual. This has been noted by the designer of the very popular game *The Sims*. Hoping to find information that would help design realistic human behaviours for his game, he found that the psychological literature provided few leads (Harris, 2003). In some cases, the design of agents design may even be based on assumptions about individual processes that we know to be wrong, for example, an assumption that individuals will exhibit rational self-interested behaviour that has been applied to some economics modelling.

A more difficult issue comes when trying to navigate the space between the levels of analysis. Doing so encounters fundamental differences in the theoretical stance to analysis adopted in the social sciences. This is particularly apparent in sociology, where the literature identifies two extreme stances. These concern the related issues of the emergence of social structures, and the nature of the individual's interactions with these social structures (Archer, 2000; Elder-Vass, 2007; King, 1999). At one extreme, individualists might hold that processes within individuals generate and drive all social structures, social facts and social actions. Each of these is reducible to features located within individuals and, at their most reduced, none exist in isolation from individuals. While this stance is identified as individualist, it is less clear which theorists, if any, are particularly associated with this stance (King, 1999), at least in its extreme form. At the other extreme, exemplified by Durkheim (Lehmann, 1993), is a stance that individual activity is driven by structural elements and the social structures in which the individuals are located. From this position, attempting to understand individuals' actions in terms of their internal psychological processes misses the more powerfully determining forces of social structures on the individual.

The bottom up nature of agent-based model construction means that its use carries with it an implied, but often unstated, theoretical stance: the model is fundamentally individualistic (O'Sullivan & Haklay, 2000). The form of the model, in which interacting individuals are represented as a system of agents, assumes that the essential components of the system can adequately be represented through the individual components (O'Sullivan & Haklay, 2000). That said, from the perspective of agent-based modelling, it is not entirely clear that the opposing camps of individualism and structuralism are as different as they would seem. Individualism exists across a spectrum ranging from methodological individualists, who hold that the understanding of social phenomena *should* be in terms of the individual, to ontological individualists, who hold that social phenomena are caused by individuals, but who do not require that understanding has to be in terms of reduction to the level of individuals (Udehn, 2002). This version of the individualist stance is implicit when the position is taken that structure is emergent, as it is held that emergent features do arise from the individuals (King, 2007). A major point of difference comes in relation to the nature of social structures.

Structuralists hold that some important features of social systems are emergent, including people, structures, and culture (Archer, 2000). Here, again, we meet a difference in meaning: this time a difference in what is meant by emergence. Just as there are differences the various forms of individualism in the debate between structuralists and individualists, there are differences in what people hold as the tenets of the various forms of emergence. In the above context, structuralists maintain that social structures are emergent features that are not further reducible to individual constituent elements. Water is cited as a physical example of the form of emergence meant here (Elder-Vass, 2007), as the properties of water cannot be reduced to an explanation in terms of the properties of its constituent elements, oxygen and hydrogen.

A second form of emergence is the appearance of patterns and unexpected outcomes from complex systems. This form of emergence takes various forms: the behaviour of a robotic device that is not explicitly programmed, but arises as a consequence of the programme interacting with the robot's mechanism and the environment (Braitenberg, 1984) is one example, while the dynamics of a traffic jam, located some distance upstream from the original obstruction, and remaining long after the obstruction has cleared is another example of an emergent phenomenon that we can observe frequently.

Returning to the water example, the properties of water can be explained in terms of the properties of hydrogen and oxygen ions, and their geometry when combined in water molecules. Thus an explanation of the physical properties of water does depend on an explanation in terms of both properties of the constituent atoms, and of the bonds between them. Similarly, emergent social properties arise, from individuals, and from the interactions between individuals.

One objection that structuralists have to individualist approaches is that they tend to include only the individuals that are present in the population. But social structures and cultural features have also developed in history, so the structures tend to incorporate elements that are not explainable in terms of individuals that exist in the population(Elder-Vass, 2007; O'Sullivan & Haklay, 2000). Individualists (King, 1999) respond that this there is no particular reason that an explanation should be limited to individuals who remain in the population

In addition to a temporal element, there is also an element of directionality in this conception of emergent properties. While a property might emerge from a group of individuals interacting, this does not mean that it is possible to take analysis in the reverse direction. That is, we usually cannot conceptually reduce an emergent properties into constituent individual components.

Finally, an individualistic formulation of social phenomena does not have to be restricted to features that exist in the individuals in perfect isolation. Rather, it might well include the relationships between individuals, their interactions, and the effects of the presence of others on their individual behaviours. These are the very materials of social psychology.

The evolution in agent-based models operates in one direction. Social structures can emerge from individual agents, but then do not feed back to affect the individual agents. This is not entirely due to the model deliberately excluding influence in this direction. Rather it is a consequence of the difficulty of the task of automatically detecting any structures that do emerge. Incorporating structures, whether emergent or existing independently of the agents, depends on them being identified and explicitly written into the model.

Agents, individuals, and the role of the environment

Agent-based models are assembled from a set of individual agents. In a modelling task, those agents are selected and designed to represent something. One obvious thing that individual agents might be used to represent is that they might represent individual people. Some of the concerns in the social sciences about carrying out analysis at the level of the individual have been outlined above. This discussion about whether it is more appropriate to carry out analysis at the level of individuals, or at the level of social structures has primarily arisen from within the discipline of sociology, which inherently straddles the transition between these different levels of analysis.

Analysis at the level of individuals, as opposed to the level of social structures, is a less controversial approach in a social psychological setting than it is in sociology. Nevertheless, the explicit representation of individuals and the interactions between them within an agent-based model can come packaged with assumptions about individuals that might be of greater concern to psychologists. The definition of agent from computer science, including the definition applying to agent-based modelling, is in terms of individual elements acting independently, in their own interest.

The word *agent* also carries a meaning within social psychology. Agents are entities that have agency: the ability to determine their own actions, and, through these, to attempt to influence their environment. Bandura (2001) defines an agent in terms of intentional action, stating that someone acts as an agent when they "intentionally make things happen by ... [their] actions" (Bandura, 2001). By this definition, *intention* forms a key element of Bandura's conception of agency.

Intentions in agent reasoning

In his discussion of intention as a component of agency, Bandura (2001) cites Bratman's (1987) model that suggests that people generate planned action by following up their desires and beliefs, with an intention to act. Under this model, planning proceeds as a series of partial plans, by which reasoning about beliefs and desires generates intentions that may progress to actual action.

This model of beliefs and desires forming the basis for planned action is, in turn, derived from earlier explanations about belief and desires, as precursors to action. These ideas are described as folk psychology (Malle & Knobe, 1997; Sutton & McClure, 2001), or commonsense understandings (Bratman, 1987), as they refer to people's everyday explanations of what lies behind another's actions (Kukla & Walmsley, 2006) in terms of informally defined mental states like beliefs and desires (Stich & Ravenscroft, 1994). Bandura (2001) maintains that this formulation is insufficient, and that human agency also depends on self-directedness to make the move from intent to action, and on monitoring and evaluation of our performance so that we can use experience to adapt our behaviour.

A link between belief and intention is made elsewhere in social psychology in the theory of reasoned action (Fishbein & Ajzen, 1975) and the subsequent theory of planned behaviour (Ajzen, 1985). In these, the link between beliefs and intentions is modified by attitudes. These theories link the intention to follow a particular course of action and the subsequent behaviour, although the construct of intention in this model is somewhat circularly described as a "subjective probability that ... [a person] will perform some behavior" (Fishbein & Ajzen, 1975, p. 288).

A third branch of social psychological thought in which the link between beliefs, desires, and intentions is raised is in the informal everyday explanations of the actions of other people. The importance of intention in these everyday explanations can be illustrated by the use of attributed intention as one of the elements in determining culpability under the law (Malle & Knobe, 1997; Kukla & Walmsley, 2006). For example, New Zealand law distinguishes between common assault (*Crimes Act*, 1961a) and assault with intent to injure (*Crimes Act*, 1961b). Unlike the definition of intention used in the theory of reasoned action, intention as used in attribution is inherently social; intent may be drawn upon by one person to explain another person's actions. Malle and Knobe (1997) found that when asked whether an action was intentional people drew on a common set of criteria. They made attributions of intention when they could infer that a person desires a particular outcome, that the person has beliefs about the actions needed to obtain it, that the person has an awareness of carrying out the action, and that the person has had the skill to carry out the action successfully. Their research indicated that desire and belief are necessary, but not sufficient for, inferring intention. Beliefs are therefore linked with intentions in people's attributions about others' behaviour. This places desires, beliefs, and intentions as resources that are drawn on in informal explanations of observed actions.

Belief Desire Intention (BDI) model

The linking of desires, beliefs, and intentions is not just explanatory. The Belief, Desire, Intention (BDI) model is proposed as the ideal theoretical basis for agent devices by some agent theorists (Georgeff, Pell, Pollack, Tambe, & Wooldridge, 1999), and has become the dominant model cited by those working in multiagent modelling. In this model, *beliefs* are a representation of knowledge about the world, which is possibly incomplete. *Desires* are a representation of the target outcome or goal. Beliefs and desires are used to develop *intentions*, identifying the means to work toward the desired goal, and eventually a *plan* of action (Georgeff et al., 1999). The BDI model, including the name, is based directly on Bratman's (1987) ideas about how people make and execute plans (Bratman, 1987), through making, executing, and reviewing partial plans.

It has been suggested that the relationship between Bratman's analysis and multiagent systems constructed using BDI is less close than the direct application suggested in the multiagent systems literature (Edmonds, 2002). In part this is because the tasks of making, reviewing, and remaking plans is proving to be a difficult technical challenge for artificial intelligence. While the implementation of BDI in multiagent systems is as yet limited, work to develop packages and standards for agent ontologies, and the languages and support for BDI logics are developing (Rao, 1996).

For social psychologists, models based exclusively on beliefs, desires, intentions, and resultant plans are unlikely to capture sufficiently some of the more important elements of social cognition. Searle (2001) offers a wider perspective on reasoning, that makes use of intentional states: mental states that are directed toward some proposition in the world. Desires, beliefs, and intentions are examples of intentional states, but these are not privileged alongside other important intentional states, such as emotional states (Searle, 2001). Although not explicitly included in Searle's list, attitudes might reasonably be added as another intentional state of interest in social psychology. An intention to act flows on from a set of intentional states, some of which may not be satisfied, giving intention a different status to the other intentional states, as it is generated from reasoning about these.

Even this extended form of cognitive model lacks some of the things that we may expect of socially located thinking. As Edmonds (2002) has noted, the BDI model for agent architecture is one model, based on the ideas of one philosopher. But other agent models exist that incorporate both rational and reactive processes, and a layer of reflective supervision of the performance of these (A. Sloman, 1998). Other architectures have been proposed that incorporate emotional dimensions such as arousal and pleasure as regulatory systems (Fromm, 2002).

Human agents, agency, and social structures

The individual agent is the basic modelling unit in an agent-based model. Within the model, each agent functions independently, and is able to interact with others. Ideally, an agent should have some ability to identify a preferred outcome, and to take action to secure that outcome (Wooldridge, 2002). In short, agents should exhibit some degree of agency: they should take intentional and autonomous action to make things happen in their environment (Bandura, 2001).

In a modelling context, agents may be used to represent individuals. If agents exhibit agency, there is an implicit assumption that we believe that humans act as agents. While an assumption of human agency might seem almost an article of faith for psychologists, there is controversy as to the degree to which individuals are able to act with complete independence. The alternative extreme to saying that the world should be understood only in terms of the actions of the agents is to say that individuals are so embedded in social structures that their actions are forced by those structures. This is a more sociological than psychological perspective. These issues about agency need at least some consideration in this thesis, as they particularly relate to the relative importance of individual agency and social structural elements in generating behaviours.

Both individual behaviours and these higher level structural mechanisms lie within the scope of analysis of agent-based models. The type of phenomena that are addressed with agent-based models are those that emerge from a number of individuals acting independently. Within a model, any structures that form an influential part of the target environment should also be addressed in an agent-based model. As the model contains both individual and environmental components, any concerns about the relative roles of individual and structural elements extend to the validity of agent-models in modelling these phenomena.

Individual and structural conceptualisations can, at their most extreme, collapse to a dichotomy in which either the individual is all, or society is all (Archer, 2000). Carried to such an extreme, all phenomena might be explained as structurally determined, with individuals dismissed as mere products of these structural forces. In such an extreme privileging of societal analysis, we miss important features of individual activity. Individuals do try to assert some influence to produce preferred outcomes. In doing so, their actions may be directed toward any component of their environment, including the natural and physical world, as well as the social world (Archer, 2000, p. 254). Sometimes these actions are highly effective, and sometimes they are highly ineffective, or even produce unforeseen and unwanted effects.

Individual conceptualisations can also be developed to an extreme, whereby all is explained in terms of individual behaviour, at the most extreme reduction, brain and computational structures (Bandura, 2001). A strictly individual perspective insists that each individual is entirely in control of his or her own destiny, and his or her actions will successfully surmount any environmental constraint. This view is essentially anarchic, as individuals will act rationally in pursuit of their own interests.

The consequences of this dichotomy between individual and structural approaches are extensive for researchers, as they determine boundaries in our framework for understanding phenomena. For example, choosing either an individual or a structuralist approach forces the researchers hand in choosing "personal agency versus social structure, self-centred agency versus communality, and individualism versus collectivism" (Bandura, 2001, p. 14).

Human agents in social psychology

Four features of human agency have been identified by Bandura (1999). Firstly, human agents have forethought that allows possible courses of action to be identified, their likely consequences evaluated and a plan developed. Secondly, they can motivate themselves to take action or to suppress action. Thirdly, they can carry out actions intentionally. After action has been taken and outcome obtained, human agents evaluate their own performance, and use the insight gained to modify future behaviour. The BDI model is consistent with three of these four features of human agency. Although not explicitly stated, the evaluative component is not inconsistent with the BDI model outlined above, although adaptive and learning systems are more demanding both computationally and technically. In summary, agency in artificial intelligence systems, specifically agent-based models, are not inconsistent with psychological views of human agency.

Many psychological theories are concerned with the individual, but are not necessarily agentic. For example, neither behavioural nor computational explanations of human activity suggest that responses to particular stimuli might involve preferred outcomes or deliberative action. Social cognitive theory is explicitly agentic, as it posits humans as active agents in their lives (Bandura, 1999). Archer (2000) notes that humans exert action in a number of distinct environments: a natural, a physical, and a social world. Bandura (1999) extends this, so that human action includes acting on the environment in agentic ways. He further notes that there are imposed, selected, and environments: some elements of our environments are uncontrollable by individuals, but humans have some freedom to choose their environment by relocation, and they can modify their environment through constructing it. Bandura (1999, 2001) states that social cognitive theory disputes a dichotomous choice between structure and agency, rather he suggests that a triad of factors are interlinked with each other:

1. internal cognitive, affective and biological events,
2. behavioural patterns, and
3. environment.

Up to now, the discussion has been based on a broad conception of agency. An agent has a preferred outcome, and acts with the purpose of achieving that outcome. Agency is about action, rather than about outcome, so an agent may or may not succeed in achieving its aims. Although actions may produce outcomes, the actual outcome may depend on other things in the environment. The actions taken by agents are purposeful.

Social cognitive theory as a candidate for research using agents

Social cognitive theory extends beyond the individual, as it suggests that the individual interacts with the environment. Neither social cognitive theory, nor artificial intelligence theories, nor most agent models specify how we might represent the environment. One result is that agent-based models tend to be focused on modelling in terms of individual agents and, as a result, inherit an exclusively individualist formulation by default. This individual formulation remains unacknowledged (O'Sullivan & Haklay, 2000). In some circumstances it may be entirely valid, and in others not. For example, a model that demonstrates how reciprocity norms can evolve (Axelrod, 1997b) may not need a representation of an informal enforcement structure, while a model that demonstrates how behaviour might evolve following the breakdown of norm enforcement may well need explicit representation of a norm-enforcing social structure.

While Bandura's (1986) formulation of social cognitive theory in terms of a triadic reciprocal causation has been around for some time, it seems to have been used very little to inform psychological theorising. While this work has been cited widely, most draw most strongly on ideas relating agency to self-efficacy, rather than the broader issues of human agency and the interactions of person, behaviour, and the environment. Removed from the systems perspective self-efficacy is a far more accessible construct, which may partly explain its popularity as an isolated construct, and Bandura himself emphasises self-efficacy as the key determinant of human agency (Bandura, 2001). But doing so strips a big idea, that of social psychology as a dynamic system in which person, behaviour, and environment interact, down to a static concept.

The absence of work calling on the more dynamic aspects of social cognitive theory does not appear to be because this part of Bandura's work is disputed. On the contrary, it is sufficiently orthodox to be presented as a theory of personality in elementary textbooks. It may be that the inherent dynamism of the model makes it difficult to incorporate into the static models that dominate psychological theory. In contrast, self-efficacy lends itself to static measures, and fits readily into static theoretical models, and may as a result be more accessible to statistical research techniques. Accessing the more dynamic aspects of social cognitive theory may require techniques that can address dynamic processes and the emergence of patterns and interactions.

Two aspects of social cognitive theory were used in the models in this thesis. Firstly, agent-based modelling, through its very nature, allows us to represent the characteristics of individuals, their behaviours, and their interactions with other individuals. It can allow modelling agents that are actively interacting with their environment. Agent-based modelling provides a means by which we might access triadic reciprocal causation directly. Secondly, the trading model is based on observational learning, as the agents adopt elements of the strategies of more successful agents.

Summary

The first five chapters of this thesis have introduced a form of modelling, agent-based modelling that offers promise in researching some social phenomena, specifically those that involve a large number of people interacting in nonlinear ways. Agent-based modelling has some implications for how we interact with theory. Further, representing people as agents carries with it some assumptions about how individuals operate. Those assumptions may fit well with some more general theoretical ideas, for example they fit social cognitive theory particularly well, offering a different way of doing research in this theoretical framework. In the next chapter, I turn to the phenomena that I wanted to apply agent-based modelling to; how trust works within networks of interacting people. The next chapter is a literature review, largely concerned with reviewing trust as a construct. The following chapter carries on to describe the ways that research into trust has been carried out, and describing the my construction of two different agent-based models of trust.

Drawing on psychological theory for building agent-based trust models

When I was in the early stages of constructing a model of trust, I was concerned that the design of the agents should be based on psychological theory. From that starting point, it seemed to me that there were two distinct forms of psychological theory that might be used. Firstly, there are theories that are entirely specific to trust, that is theories and algorithms about who trusts, and how they trust.

Secondly, there are the more general theories from social and cognitive psychology that might be applied to trust. From social psychology we have theories that, in a specific social situation, people will, on average, make particular assessments of others, and behave in particular ways. And from cognitive psychology and cognitive science there are theories about how people gather information and make decisions.

Whether a specific theory about how trust functions or a more general psychological theory is used, we encounter another difficulty in incorporating psychological research into an agent-based model. Despite the vast body of psychological knowledge, findings, and devices, much psychological theory is not in a form that makes it suitable for direct incorporation into agent-based models of trust. There are relatively few theoretical ideas, in the form that they exist in psychology, that can be applied directly in designing individual agents that are well supported by theory. For example, a finding that higher educated people are, on average, more trusting might be reported in terms of means and standard deviations, and results from testing hypotheses about these means and standard deviations. In some circumstances, applying a simple mean to all members of a population might be a reasonable representation. At other times, the modeller is likely to want to represent diversity in the population in a credible way. It can be difficult to extract a supportable model of that diversity. While analyses are likely to have been carried out on the assumption that the distribution was normal, often the only diversity information available is in reported standard deviations. The shape of the distribution, or even the range of values found, is often not reported. Even when using psychological theory that is well supported, the modeller is unlikely to have sufficient information to be able to incorporate theory directly into the model. This means that the modeller will, almost certainly, have to use judgement at some point in the modelling process.

A related issue arises in deciding what inputs are necessary and sufficient to produce a reasonable model. This is related to another problem that is recognised as extremely difficult for engineers wanting to design computers with autonomy. How can we predict what components will be necessary and sufficient to produce a particular desired outcome? Similarly, in social science research, we would like to have some confidence that a model included all of the essential components that go into generating a particular high level behaviour. One possible answer comes from Wolfram's (2002) work in cellular automata. Once sufficient complexity is added to generate the first instance of complexity, adding more complexity does not produce any different outcomes. Or as he puts it, simple rules may be sufficient.

Chapter 6

Trust

The foregoing has set out some of the theoretical issues about agent-based models as a research method. In this chapter I review the literature on trust as a construct, and as it might be relevant to developing agent-based models of trust. The two chapters following this one describe the development of two models of trust within a network of individuals and, in the process, explores some of the more practical issues arising from carrying out research using agent-based models.

Much of the research and analysis into what trust is and how it works is concerned with trust as an individual level phenomenon, but there are also situations in which features arise that may be founded in trust operating in populations. For example, there is a relationship between mean levels of generalised trust reported in national surveys and the population level outcomes such as economic productivity and health outcomes. In this situation, data at one level of aggregation seem to be driving outcomes at another level of aggregation. One mechanism proposed as possibly linking these is that the economic costs and benefits of trust and distrust are important in enabling trade and cooperation and disabling fraud.

There are some aspects of trust that may be relatively static. For example, people may have varying degrees of propensity to trust others that are sufficiently static to be considered a personality characteristic. Other aspects of trust are more dynamic: decisions to trust or not are made frequently; states of trust may or may not exist at any moment; and these states may change rapidly. The dynamics within networks of people interacting in ways that engage trust may also be important in determining the results of a system's evolution; determining the attractors to which a system may be drawn. These dynamics might lead to systems of individuals adopting a culture of trust or distrust through the dynamics of their individual level processes and interactions.

The previous chapters have proposed agent-based modelling as a method that can be used to develop understanding of systems that involve dynamics and aggregation. Constructing such a model depends on understanding and representing the processes of trust at an individual level. This chapter begins by reviewing the literature in this area.

Why be interested in trust?

Trust is a pervasive feature of our daily life: individuals, businesses, organisations and governments routinely engage trust in their everyday interactions. The variety of entities relying on trust, the variety of relationships in which trust appears, and the different circumstances in which it is invoked, is reflected in the richness of form and detail of the ideas surrounding trust. For example, trust in an intimate relationship takes a different form, and has entirely different boundaries, to trust between a government and an individual citizen.

Trust is important at an individual level, because most social relationships call on some degree of trust. This is supported by studies that correlate factors associated with trust and social relationship factors. For example, high levels of trust are related to satisfaction and commitment in relationships with partners (Couch, Adams, & Jones, 1996; Couch & Jones, 1997) and high levels of mistrust are associated with interpersonal problems (Gurtman, 1992). An individual's self-rated propensity to trust is correlated with self-disclosure (Bierhoff, 1992), and people scoring high on trust scales report themselves as being more warm, gregarious and empathetic than do people with low scores (Couch et al., 1996), while others also think them to be more likeable (Rotter, 1980).

Trust is essential in commerce as much as it is in personal relationships: a restaurant takes an order for food without prepayment, with the expectation that at the end of a meal the diner will pay for the food ordered; employees work for a period, with the expectation that the employer will pay them at the end of that period. Arrangements in which turns are taken to meet obligations, for example the completion of a service and subsequent payment on a monthly invoice, are fundamental in normal commerce, and essential for a business's ability to function efficiently. Such arrangements depend on trust: the first person to meet their obligation expects that the second person will meet theirs.

In commerce the alternative to relying on trust is to employ active risk management strategies, whether by controlling exposure to risk by restricting the size of transactions, or by investing in a formal contract, verification and enforcement procedure. In contrast to enforced compliance, trust allows businesses to benefit from cooperation, for example through coordinated marketing efforts and joint ventures. The alternatives to trust are expensive in business, whether through lost opportunities, or through adding extra dead-weight costs. Overall, trust is recognised as an important factor in business efficiency (Lewicki, McAllister, & Bies, 1998).

The importance and variety of applications of trust has resulted in an extensive literature on how trust is cued, facilitated and threatened in different contexts and in different relationships. It is this literature that provides some guidance as to the features of trust that might be incorporated into a simulation model of trust. A subset of this literature, relevant to developing a cognitive model of trust, will be assembled in the next section.

A second set of literature on trust is concerned with generalised trust. Whereas the previous set of literature is how trust works in relationships, this second set of literature is more concerned with beliefs about trust. Generalised trust is generalised to the extent that it is not specific about who is trusting whom, with what, and in what situations. It is not about indiscriminate trusting, as this is characteristic of gullibility rather than of trust. Rather, generalised trust is a non-specific expectation about the behaviour of others, and so might be thought of a belief about norms of trustworthy behaviour.

A high expectation of trustworthiness, a high level of generalised trust, has been found to correlate with some national performance indicators. This suggests that the aggregated effects of trust and trustworthiness may also be important.

Beyond individual relationships, trust has been of interest as one of the social devices that allow the institutions of society to function (Putnam, 1995). At the societal level, trust is important in many arenas: in the development of communities; in the maintenance of order; in maintaining the legitimacy of governments; and in the efficient operation of economies. In this setting trust is identified as an element of social capital: it is one of the things, like robust social networks and widely observed social norms, that go to make a well-functioning society. The capital part of social capital recognises the economic value that can be realised through efficient communication, cooperation and interactions. Placed in a wider social environment, trust moves from being a feature of interpersonal interactions to having a broader role in the function and maintenance of society, where it is explicitly identified as a component of social capital.

The loss of social capital in modern societies has been of concern (Putnam, 1995), and the more specific role of trust as a component of social capital has been highlighted in concerns about the restrictions and costs of relying on enforcement (Knack & Zak, 2002) or family ties (Fukuyama, 1995), rather than on trust.

A measure of generalised trust is usually extrapolated from a survey question that asks whether the respondent believes that, in general, people can be trusted. As measured by this variable, generalised trust has been significantly correlated with aggregate measures of economic performance (Knack & Zak, 2002), and with participation in democracy and political stability (Inglehart, 1997). The economic and political importance of these performance outcomes for nations and development has spurred research into generalised trust by the World Bank and the OECD.

Unlike a direct understanding of the mechanisms of trust, generalised trust is not obviously related to building an agent-based model. Rather than being important in the building process itself, generalised trust is of interest because if provides a hint that trust might accumulate beyond individual interactions. Between the literature on the phenomena of trust in individual transactions, and the literature on correlations between attitudes to trust and macro-level measures is a gap extending to both the mechanism and construct. One possible bridging of this gap is the development of an understanding as to how trusting interactions might combine to produce the social environment in which individual relationships and interactions are embedded.

Previous research into trust

In the literature, the term *trust* is variously used for constructs that are, alternatively: an attitude; a virtue; a behaviour; a trait; a conscious rational decision; or the outcome of an unconscious cognitive process. This leads to a difficulty for the development of a simulation model of trust - are we to simulate an attitude, a behaviour, a personality trait, or a cognitive process? As we might expect, given the diversity of features of trust, research into trust has taken a number of approaches.

Some researchers have been concerned with the development and application of trust scales. The development of scales is useful, both in providing a measurable quantification of aspects of trust, and also in the theoretical effort that is both needed for, and enabled by, the development of scales. An early trust scale, the Interpersonal Trust Scale (IPT) (Rotter, 1967), provides an example of this. This scale actually accesses generalised trust: scale questions ask respondents to identify the outcome that they would expect in various relatively non-specific situations. Another, later, reported trust scale (Couch et al., 1996) has three subscales; generalised trust; personal trust within close relationships; and something that the authors describe as network trust. Rotter (1980) explicitly states that the propensity to trust, which we might presume from his research to mean the propensity for generalised trust, is a relatively stable personality trait. Scale development inherently takes the position that the degree to which individuals tend to trust is a measurable individual characteristic (Goto, 1996).

Scales have been used in correlational studies. For example, Rotter (1980) applied his scale to studies investigating whether trust was simply another name for gullibility. He was able to report (Rotter, 1980) that high trust did not appear to be simple gullibility. He notes that the difference possibly lies with the use of evidence; gullibility is believing despite being provided evidence that suggests that may be unwise, trust is believing in the absence of evidence to suggest that this is unwise.

Measures of trust are also used in survey investigations into the relationships between social, economic, and population indicators. In a series international surveys, Inglehart (1997) found that generalised trust was accompanied by reported subjective well-being, political and organisational participation, lower levels of income inequality, and low levels of extremism. In these surveys, the measure of trust is a single question that is very similar to, but less specific than, some of the questions asked in the IPT scale. Zak and Knack (2001) used earlier versions of these surveys to generate a model predicting mean trust levels in a population using income per head, education, corruption, and income inequality. They suggested that levels of generalised trust may have a causal role in determining economic performance, through the higher costs due to increasing diligence efforts and the restriction of economic interaction where diligence measures are not cost effective.

The factors associated with trust have also been explored in experiments using vignettes. These have been useful in identifying the sorts of information that people use in making trust decisions. For example, Yamagishi (2001) extended Rotter's (1980) finding that trust is distinct from gullibility with a study that used vignettes. The situations described in the vignettes were identical, but with added a) two pieces of positive information, b) one piece of positive information, c) no added information, d) one piece of negative information, or e) two pieces of negative information about a person to be trusted. High trusters were particularly sensitive to negative information, being less likely than low trusters to trust someone once they were given two additional pieces of negative information. As gullibility is a decision to trust despite the presence of negative information, this suggests that high trusters may be *less* gullible than low trusters. Rather, as Rotter (1980) suggested, they seem to make better use of the available information.

Vignettes have been used to investigate other key factors in trust, the closeness of the relationship and the degree of risk involved. Goto (1996) also used a trust scale to identify high and low trust participants in an experiment to test the involvement of two other factors that are often identified as important in discussions about trust: the closeness of the relationship and the degree of risk involved. She found main effects for all three variables. As might be expected, people were more likely to trust if they were high trusters, if the relationship was close, and if the stakes were lower. There was also an interaction effect, such that people were less likely to trust persons with whom there was a closer relationship when the stakes were high, and more likely to trust strangers when the stakes were low.

Other researchers, beginning with Deutsch (1958), have used formal games with simple rules to carry out experiments in trust. Formal games have very simple moves, and a restricted set of possible moves. This allows the game to be analysed formally, and also provides a useful experimental device for investigating strategic behaviours. In the early stages of applying game theory to trust research, Deutsch used the Prisoner's Dilemma in his experiments. In the Prisoner's Dilemma, the best individual outcome arises when a player defaults, and his opponent does not. No matter which way the opponent plays, defaulting produces the best individual outcome, and so this strategy is regarded as a stable strategy. Defaulting does not, however, produce the overall optimum outcome, as this is achieved when both players cooperate.

At the time of Deutsch's experiments, few formal games had been defined. Subsequently it was recognised that, while trust may be a component of a player's decision in the Prisoner's Dilemma, this task more particularly concerns cooperation, and the breakdown of cooperation. Furthermore, there is no explicit agreement between the players. Rather, cooperation is implicitly involved in a preference for the optimal outcome. Later in the development of formal games, games that more specifically target trust were developed (Camerer, 2003a). Unlike the Prisoner's Dilemma, trust games are turn-based games between players with different roles. In trust games, the first player makes a decision whether or not to trust the second player. Thus only the first player is playing a trust game. The second player then decides whether or not to honour that trust. These games have provided an experimental device for running experiments in trust. For example, Yamagishi and colleagues (Yamagishi, 2001; Kashima, McKintyre, & Clifford, 1998) have used trust games for investigating cultural differences in trust between Japanese and American students and their respective expectations about norm enforcement.

Methods used in trust research

While they might be reasonably uniform throughout a population, understandings of trust are not universal. Cross-national differences in the endorsement of belief that people can be trusted (Inglehart, 1997) and differences in the application of trust (Yamagishi, Cook, & Watanabe, 1998; Reeves-Ellington, 2004) suggest that common understandings and behavioural norms around trustworthiness and trust are not arrived at as a matter of course. Rather, it appears that the specific expression of trust has a cultural component. An example of the variety of trust behaviours is provided by the studies of American and Japanese students by Yamagishi and colleagues (1998). Removed from a cultural environment where informal enforcement of social norms is very strong, to an experimental environment where there is no sanction against defaulting, Japanese students are less trusting and less trustworthy than US students. This extends to the function of trust in natural settings, where differences in trust cultures can be one element in a mixture of cultural incompatibilities within an organisation, such as between Bulgarian and American staff at a university in Bulgaria (Reeves-Ellington, 2004). The differing expressions of trust in different cultures would suggest that trust and trustworthiness may have developed differently in different social environments.

The existence of different trust cultures tends to suggest that trust may develop differently in different environments, rather than being a universal norm. Although there is some suggestion that there may be a genetic propensity for different degrees of trust (Zak, 2003), it is also likely that the social environment determines the development and functioning of trust in individuals within a society. This scope for the development of trust, raises the possibility that trust might be influenced or modified by social processes. Combined with the relationship of trust between a number of important indicators of societal performance, an understanding of the processes surrounding the development and loss of trust among the members of society is potentially important across a range of social issues. Investigation of change in a large social environment, where the environment is itself determined by individual actions, is inherently an investigation of a dynamic system. This tends to suggest that dynamic analyses may also be useful in investigating trust.

Many studies of trust are inherently static, placing trust in relation to other variables. Formal game experiments may also be of a static nature, as a single round of a game is a one-off experiment. This changes when a game is repeated, as the players have an opportunity to take risks in order to demonstrate that they are prepared to cooperate to reap larger combined benefits. Repetitions of the game or, more importantly, expected repetitions may change things substantially, and may add a dynamic element.

How does trust manifest?

In order to model trust, we need an understanding of how and when a person trusts. One suggestion (Castelfranchi & Falcone, 2000) is that trusting is a three-stage process. This process consists an evaluation of the other person, a decision to trust, and acting on that decision, and this process depends on the specifics of person and situation. These are individual cognitive differences in attitudes and in the strategies employed in the assessment process.

Besides these differences in the cognitive processes of trust, there are other sources of the individual differences in people's propensity to trust. There are differences correlated with demographic factors, with learned experiences, and with personality traits.

Cognitive approaches to trust

Castelfranchi and Falcone (2000) agree with Hardin (2002) that trust consists only the assessment and decision stages of the three stage process, and that trust does not extend to the subsequent actions that may depend on trust being established. While Hardin(2002) includes both assessment and decision-making as the components of trust, he concentrates almost all of his subsequent analysis on the decision-making step. This concentration on the decision-making component of trust is particularly characteristic of economic and game theoretic models of trust.

Game studies, in particular, often short-circuit the assessment phase almost entirely, as experimental controls often restrict participants from making an assessment of their opponent by restricting exactly the sorts of information people might draw on to make their assessments. Typically, players cannot draw on information about their opponent, as opponents are anonymous, unseen, and unheard. In the extreme, although players may have been told that they have human opponents, the opponents may be a computer. Approaching trust as a decision-making task tends to result in a model of trust that is quite calculative. The result is that these models do not seem to capture much of the character of trust, which is not experienced as a calculative process. Notably, trust also has an emotional dimension that is missing in these analyses, that may be inconsistent with calculativeness (Williamson, 1993).

In contrast to game theoretic approaches to trust, philosophers writing about trust put assessment in a much more prominent position, relocating trust as an attitude (Holton, 1994; Lahno, 2001) that may affect our decision-making, rather than just the decision-making process itself. Explicit consideration of the assessment process allows us to incorporate an important feature of trust that is not encompassed in game theoretic accounts of trust. Framing trust as an attitude allows us to suggest that beliefs, emotions and actions might all be important in trust.

Considering trust in terms of assessment and decision making gives us one perspective, of trust as a cognitive process. But other factors come into trust, notably those associated with individual differences in the propensity to trust. Agent-based modelling allows us to represent individuals as batches of identical average individuals, but it also allows us to extend the representation of individuals. We can assign each individual agent a unique set of characteristics; agent-based modelling allows us to represent individual differences explicitly. We can either use a population of identical individuals, or we can assign characteristics to each individual agent, or to groups of agents. To do this, we need some understanding about how trust assessments and decisions are influenced by individual differences, and how these differences are distributed. In addition to cognitive models of trust, some demographic characteristics have been correlated with trust, and dispositional measures of trusting. We can incorporate these factors, and others such a group membership, explicitly, and we can allow the factors to be modified.

Trust as rational decision-making

If we accept the view that trust is a process of assessment and decision-making, it follows that trust is a cognitive process. The nature of that cognitive process is, however, less clear. The usual view of trust is that it is necessarily a rational cognitive process (Hardin, 2002). While Hardin does not define rationality, his arguments are consistent with assumption that he means instrumental rationality. Instrumental (or economic) rationality is based on a core assumption that people have a set of identified preferences about outcomes, that they can rank these preferences consistently, and that they will actively attempt to achieve their most preferred outcome (Colman, 2003). A model of trust as an instrumentally rational process might take this form: a person assesses a situation and the intentions of another person (Hardin, 2002), identifying possible outcomes and their value and likelihood. On the basis of these assessments the person decides to trust or not trust depending on the assessed likely outcome.

This view of the cognitive process associated with trust is consistent with an assumption that individuals will attempt to optimise their individual position. In economics this individual position, the sum value to the individual, is called utility. Economists have developed mathematical descriptions of many situations that may involve utility maximisation. According to a utility maximising view, an individual who successfully maximises individual utility is exhibiting rational behaviour. In turn, successful optimisation may depend on a mathematical or logical solution to a formal statement of the problem, and a response consistent with this solution may be the only response that would be regarded as rational.

One form of situation that economists have explored is the theory of formal games. These are very simple games that provide an opportunity to carry out experiments in strategic thinking. The very simplicity of the game rules limits the possible moves, and restricts the sources of information available to players. While formal games are simple, there are many ways that game rules can be varied, and the game conditions can be manipulated, to allow them to be used as experimental devices.

Formal games are particularly useful for exploring how people approach tasks that call for competition or cooperation, and for investigating how people seem to draw on unwritten rules of trust, trustworthiness, and fairness in these tasks. While the strategies being employed cannot be observed directly, researchers infer the strategies that are being employed from the observed choices that players make when playing formal games. Depending on the specific rules and conditions these strategies might, for example, include: bluff and deception; trusting other players; insisting on fairness; attempting to negotiate agreements; and punishing defaulters.

Optimal solutions exist for many formal games, and often solutions amount to a recommendation that individual players should always default (Camerer, 2003b). Default often makes for a successful individual, one-off, strategy because it minimises the potential damage arising if an opponent defects, and seizes any opportunities that present if other players do expose themselves to exploitation. In these analyses of optimal solutions, attempts at cooperation are seen as irrational strategies, because they provide openings for exploitation.

According to this analysis, if players played strictly rationally, in terms of seeking an individually optimal result, cooperation would never arise. Players would never take the risks necessary for cooperation, and any player who did take such a risk would be exploited. According to the formal solutions, trust and cooperation are not rational strategies, at least to the extent that rationality means utility maximisation.

Trust as reasoned, but not rational, decision-making

While mathematically optimal solutions do exist for games, the evidence from game experiments suggests that people rarely use these strategies. Rather, they tend to find an unexpectedly high proportion of people using non-optimal strategies (Camerer, 2003a). This finding holds even when experiments are carried out using business students, who should be familiar with the optimal strategies identified by game theory (Camerer, 2003b; McCabe, Smith, & LePore, 2000). Economists assume that the use of optimal strategies demonstrates strategic sophistication on the part of the player, while other strategies do not (Costa-Gomes, Crawford, & Broseta, 2001). Selection of other than formally correct solutions is sometimes assumed to demonstrate flaws in human reasoning and a failure of rational thinking.

The frequent use of non-optimal strategies rather than formally correct, optimal, rational solutions is reminiscent of patterns of decision-making that have been explored by cognitive psychologists. For some particular types of problem people tend to choose options that are inconsistent with formally correct solutions. For example, people tend to make systematic errors with probabilistic data, regardless of their formal training in mathematics and statistics (Stanovich & West, 2002).

There have been a number of alternate interpretations as to why people may make apparently incorrect decisions (Stanovich & West, 2002). These include:

1. People may attempt to make rational decisions, but the limited computational resources of the brain and limited availability of information from the environment lead to errors (Simon, 1978; Selten, 2001).
2. People may attempt to make rational choices, but using a form of statistical reasoning that draws on their existing knowledge to provide prior probabilities (McKenzie, 2003).
3. People may make mistakes with artificial problems because we are optimised for a different type of problem (Cosmides & Tooby, 1997; Gigerenzer, 2000) or because the tasks as presented are framed differently to natural tasks. For example, errors made when tasks are framed in terms of probabilities disappear when data are presented as frequencies rather than probabilities (Gigerenzer, 2000).
4. Researchers may be making a false assumption about how participant interpret their experimental tasks, to the extent that participants may be solving an entirely different problem to that envisaged by researchers.

The last of these is particularly relevant to tasks that may involve cooperation. Formal solutions are derived based on what are the assumed preferred outcomes. But it may be that game participants' preferred outcomes are fundamentally different from these assumed preferences. Specifically, formal solutions for games usually assume that the preferred outcome gives the short-term maximum value for the individual, while people may actually prefer an outcome that gives the maximum value for the whole group, or an outcome that gives maximum utility consistent with maintaining relationships. If this is so, it may be that people really are making rational, and correct, choices (Stanovich & West, 2002; Van Hezewijk, 2004). If the experimenter's assumptions about preferred outcomes are wrong, then their assumptions about what is the correct solution will also be wrong.

Selection of an alternate outcome might, in itself, be seen as being less than perfectly rational. Why would anyone prefer a compromise outcome when they might pursue an individually optimal outcome?

One potential reason, a reason that is entirely rational, is that the long-term application of optimal strategies results in worse outcomes for the individual than do non-optimal cooperative strategies. In traditional economic games, rational players attempt to maximise their short-term utility. Often this optimal strategy recommends defection. Further, a rational player should assume that his opponent is equally rational and knowledgeable, and will also use optimal strategies. Carried to its logical conclusion, optimising strategies reduce to a recommendation that a rational player should default, and expect his opponent to do the same. This effectively mutates the player's preferred outcome. Through individually attempting to obtain maximum gain, players end up attempting to minimise the potential losses that that might be inflicted on them. Loss minimising strategies may protect against the worst case, which is the potential damage that might be sustained as a result of a failed attempt at cooperation, but they cannot produce the benefits of a cooperative outcome. As a consequence, and depending on the situation, rational decision-making can produce a middling outcome at best, and entrenchment of a stable minimum at worst. Seen in this light, rational decision-making is not especially rational when it precludes cooperation, when cooperation may produce better outcomes both collectively and individually.

The above commentary particularly relates to a choice between cooperation and defection in games, but is as relevant to choices between trust and mistrust. While trust and cooperation are not synonymous, or mutually dependent, they are closely related concepts (Ullmann-Margalit, 2002). As with cooperation, trust offers potential benefits for the truster, but at the cost of exposure to damage if the trustee defaults. The concepts are so close that the incompatibility of instrumental rationality and cooperation invites a challenge to assumptions of trust as a necessarily rational process. At its most stark, trust simply cannot function as an economically rational process, for the same reasons that cooperation cannot be economically rational (Hollis, 1998). Further, trust is fundamentally irreconcilable with economic rationality; optimisation processes are inherently calculative, and this is incompatible with trust (Williamson, 1993). Trust is irrational, at least to the extent that it cannot be rational in the economic sense. This does not necessarily mean that there is no reason involved in trust, but the form of reason involved may not necessarily be rationality where we take this to mean short-term utility maximisation (Hollis, 1998).

While a view of trust as an other than rational process runs counter to some of the trust literature, even economists are concerned about the strict optimisation required for economic rationality, for example, (Williamson, 1993) notes that strict rationality is "mind-boggling". Beyond being mind-boggling, strict and exclusive rationality can be crippling for decision making. People who have fully functioning rational faculties, but damage to the brain structures that generate and integrate emotional responses with decision-making are able to identify solutions and calculate potential outcomes, but find it impossible to make decisions, possibly because they find it impossible to generate and identify a preference as to outcome (Damasio, 1994).

Trust as an attitude with emotional content

As noted earlier in this chapter, some writers suggest that a sequence of events surround trust: an assessment, a decision and an action contingent on that assessment and decision. The analysis has, to this point, been restricted to the decision-making phase of this process. Thus far, a cognitive formulation of trust has not encompassed assessment of the person being trusted in the context of the situation. As already noted, it is possible to consider the assessment component more explicitly. From this vantage point, some philosophers have suggested that we consider trust as an attitude (Jones, 1996; Lahno, 2001) rather than a decision following from a rational analysis.

Considering trust as an attitude allows us to draw on the social psychology of attitudes, and in doing so we gain a location for some of the features of trust that are awkwardly left over after assembly of a purely decision-making model of trust. Saying that trust is an attitude is an explicit acknowledgement that trust is an evaluation of something. In the case of trust, it is an evaluation of the person being trusted. Attitudes may be developed through cognitive mechanisms, but this is only one possible derivation. Notably, as well as a cognitive component, there is a substantial emotional component in forming attitudes, that is notably absent from an exclusively decision-making formulation of trust.

The emotional content of trust differentiates trust from a purely calculative assessment, and from reliance more generally. There is a qualitative difference between the betrayal of trust and the failure of reliability. For example, we feel more distressed by the betrayal of trust than by a breakdown in reliability. This may be because trust is an attitude is located within an interpersonal interaction, rather than being an attitude directed at some inanimate object (Holton, 1994; Lahno, 2001). When we trust someone, we make some assessment of their goodwill or shared interest, and we assume that they will select their actions based on these (Baier, 1986; Lahno, 2001). An assessment and decision to trust carries with it a signal that the other is a person (Holton, 1994) and, in turn, being trusted carries with it the value of having been recognised as a person (Lahno, 2001).

The importance of the emotional content of trust may raise concerns about attempting computer simulation of trust. The experience of emotion is one of the hard, possibly unsolvable, problems of artificial intelligence. It is certainly well beyond current artificial intelligence systems to simulate the quality of experiencing an emotion.

While simulating the quality of an emotion is (currently) impossible, this does not preclude a simulating the effects of emotion on decision-making. One likely function of emotions is that they are useful in information processing. They seem to be important in guiding our attention, and selecting which information is relevant, in apply that information, in motivation, and in selecting preferences. It may also be that emotions are also useful in deciding how much analysis to call on in solving a problem.

An applied example of the simulation of emotion is the provision of an emotion interface to artificial decision-making devices. The purpose of this work is to try to replicate the efficiency (Chown, Jones, & Henninger, 2002) and unpredictability of humans (Henninger, Jones, & Chown, 2003) in an agent-based model of a special forces unit. The system designed by Cho (2002) uses a neural network to simulate emotions. This neural network controls the decision-making system. When emotional arousal is high, through fear or confusion, this emotion simulating system overrides the rational tactical decision-making system, falling back on a simpler strategy, such as fleeing. This model is based on the idea that emotional responses may have a hand in determining the form of reasoning that is applied, implicitly presupposing two different forms of reasoning in humans, although the authors do not mention this.

Trust in a dual-process system

Researchers in cognitive psychology, however, have explicitly proposed two distinct reasoning systems. System 1 is innate and intuitive, and is shared by humans and other animals, while System 2 is more analytic, and is an exclusively human ability (Evans, 2003) which can act as a check on the innate system. These two systems have spectacularly different characteristics as listed in review articles by Kahneman (2003) and Evans (2003), and a book chapter by Sloman (2002).

System 1 is efficient in the sense that it produces a result very quickly. This may be because it applies parallel processes to information, rather than serially assembling pieces of information in a systematic way. It is automatic and effortless, and the process is unconscious, at least until the outcome reveals itself to us. This may be because System 1 is realised in neural network type circuits, which are programmed through association of input and output conditions and so are essentially atheoretic. We may be unaware of our reasoning process simply because there is no theory behind them. Nevertheless, the patterns developed in in System 1 might be modelled as heuristics: simple rules and strategies that use minimal information. The slow-learning characteristic of System 1 may also reflect the repetition required to incorporate new situations into a neural network. System 1 is also associated with the greater involvement of emotion in the decision-making process (Kahneman, 2003).

System 2 is much slower, requiring an effortful, sequential, application of rules. It is more flexible, allowing analysis through abstract reasoning and the generalised application of rules. System 2 is not associated with emotional involvement in decision-making.

Some difficulties of thinking about trust exclusively in terms of rational decision-making have been addressed earlier in this chapter. Further to this we might consider the cognitive processes associated with trust in the light of the dual-process model of reasoning. Decisions to trust are often made in exactly the circumstances in which rational decision-making is difficult, particularly in regard to speed, complexity and the degree of missing information. Situations calling for a dependence on trust can be complex, requiring rapid risk assessment and mind-reading of another party's intentions in the face of limited information. While these characteristics may freeze rational decision-making processes into indecision, the same information might be sufficient to fire up a heuristic and produce a quick result. The alternative to trust as a rational decision process is that it arises initially through a reasoned, but not rational intuitive process. Such a process might be intuitive, calling on System 1 to make assessments about whether to trust quickly, effortlessly and unconsciously, using emotional content and heuristics. While heuristics can provide a very quick and efficient and accurate assessment (Gigerenzer, Todd, & the ABC Research Group, 1999), they can also mislead (Kahneman, Slovic, & Tversky, 1982). In the two system model, the intuitive system provides a rapid decision-making process, and the rational system provides a safety override that allows us to stop and take second thought as to whether to trust.

Evolutionary games

It should be noted that the earlier discussion on game theory applies to a particular conception of game theory: economic game theory. In contrast to economic game theory, evolutionary game theories make different assumptions about the criteria for success, and about how optimal strategies might have developed. According to evolutionary game theory, a strategy is optimal if it was adaptive when it evolved. The development of characteristics through evolution tends to have been best accepted for biological structures that are adaptive in the physical environment (Cosmides & Tooby, 1992), but for social animals, it seems at least possible that adaptation to the social environment may also have been important. This is a more controversial proposition, for a number of reasons. Evolutionary psychology suggests that we have evolved particular psychological tendencies. While these tendencies might be associated with behaviours that are observable, many of the brain structures that facilitate them are not readily identifiable, and the nature and degree of specialisation proposed for such structures is controversial (Fodor, 2001).

Another reason that evolutionary psychology is more controversial than biological evolution is that there is no fossil record of behaviours. One approach is to use computer simulation to evolve behaviour, often using games to define an environment. This approach has some similarity to the economic game theory. As economic game theory investigates optimal strategies for access to constrained economic resources, evolutionary game theory investigates which strategies are sufficiently successful in natural situations.

Computer modelling experiments using this approach have demonstrated that social norms can evolve, and that this might be a source of altruism. Some of these norms are paradoxical if considered in the light of rational theory. For example, reciprocity is so universal a social norm that we might want to consider the possibility that reciprocity might be an inbuilt behaviour in humans. Computer modelling suggests that such a norm can be evolved from an iterated Prisoner's Dilemma, in which a Prisoner's Dilemma type game is played repeatedly. This allows the agents to evolve strategies (Axelrod, 1997b). The optimal solution generated by the computer model, based on an evolutionary approach, is quite different to the optimal solution produced by another technology, formal mathematics and logic.

An important difference between economic and evolutionary game theory is that economic game theory tends to produce an optimal solution, but has little to say about how this is reached. In contrast, evolutionary game theory is inherently concerned with how an outcome evolves, and so with the dynamics of that evolution. Computer models allow the exploration of the dynamics of how strategies evolve, and so can address stability as well as optimality. This can identify systems in which game theoretic optimal solutions are unreachable stable equlibria (Enquist, Arak, Ghirlanda, & Wachtmeister, 2002), or in which two non-optimal strategies may reach a stable coexistence (Abreu & Sethi, 2003).

Animals, including humans, encounter a range of situations, both physical and social. The range and unpredictability of possible situations require strategies that are sufficiently efficient, generalisable, and successful. Adaptive strategies do not have to be optimal in every situation, on every dimension. A relatively small set of highly generalisable, simple strategies may be adaptive if they can identify and implement a best guess strategy very quickly, with the expenditure of few resources. It may be that our cognitive facilities are something akin to a station-wagon design. These are not optimal for luxury, performance, handling or load capacity but are, nevertheless, a successful design because they provide a reasonable solution for a range of functions.

Table 6.1: *Sentence lengths in a typical Prisoner's Dilemma Game (player 1's sentence: player 2's sentence). For example, if player 1 cooperates and player 2 defects, then player 1 is sentenced to 5 years, and player 2 goes free.*

		Player 1	
		Defect	Cooperate
Player 2	Defect	3:3	5:0
	Cooperate	0:5	1:1

Early game experiments in trust: the Prisoner's Dilemma

The first psychological experiments in trust were carried out using formal games, usually the Prisoner's Dilemma (PD). In the PD game, two players must simultaneously decide whether or not to cooperate. If both players cooperate, both get a high pay-off (a low sentence). If one player cooperates, and the other does not the cooperating player gets a very low pay-off (a high sentence) and the defaulter gets a very high pay-off (no sentence). If neither player cooperates, both get a low pay-off (a moderate sentence). The Prisoner's Dilemma produces the greatest total outcome, the welfare maximising outcome, if both players co-operate, but the greatest individual outcome is achieved if a player defects against a cooperating partner. An example of the payoffs in a PD game is shown in Table 6.1.

In early research PD was used as a device to investigate trust, and it is sometimes still interpreted as a trust game. A player makes a cooperative move only if he both expects that the other player will attempt to cooperate, and if he decides to cooperate rather than take advantage of the other person's cooperation. While cooperation and trust are not synonymous, they are closely related (Ullmann-Margalit, 2002). Trust is an essential prerequisite to acting cooperatively in the PD. Assessing a cooperative move in the PD against a definition of trust as an "accepted vulnerability to another's possible but not expected ill will (or lack of good will)" (Baier, 1986), a cooperative moves does involve both an expectation that the other will act with good will, and an accepted vulnerability to that. But a decision to cooperate goes further, as it involves both a decision to accept risk and a decision to cooperate (Cook & Cooper, 2003).

These early researchers assumed that there was no rational optimal strategy for the PD. In the absence of an individually optimal strategy, it was assumed that the most rational strategy was for the players to try to achieve the welfare maximising outcome (Deutsch, 1958). The welfare maximising optimum requires that players both make cooperative moves. Although the PD requires both trust and cooperation, it was initially interpreted as being primarily about trust, as all players were assumed to be striving to achieve cooperation.

Other researchers were more concerned that the PD did not separate trust and cooperation. This is not entirely unexpected. The close relationship between cooperation and trust, and the use of PD in investigating trust suggests that they have to some extent been regarded as synonymous. But trust and cooperation are not the same thing. A decision to cooperate in the PD represents an initial decision to trust and a subsequent decision to cooperate. But defecting might represent either mistrust and damage limitation, or an expectation that the other will cooperate coupled with a decision to exploit this. The PD is particularly exposed to this, because both players move in a single step in this game, and so the trust step is necessarily packed with the decision to cooperate. It is difficult for researchers to convincingly separate trust from other related concepts, especially cooperation in the PD.

Investor games

The PD is only one of a number of formal games. More recently a number of other games have been devised, the designs of which access different strategies. Experiments in trust have recently been revisited by social psychologists and behavioural economists exploring a number of related concepts: exclusive altruism with no cooperation component; expectations of fairness; reciprocity; trust; and the enforcement of norms. This research has used a newer set of games, classed as trust games (Camerer, 2003b), that access trust. In these games, trust decisions are split from cooperation decisions. This is largely achieved by manipulating the timing of moves made by the players. In the PD both players move only once, and that move is made simultaneously by both players.

One such trust game is the Investor Game (Camerer, 2003a; Rieskamp, 2001). In this game, one player takes the role of Investor and the other player takes the role of Borrower. The Investor is given an allocation of money to invest, and must decide how much to invest with the Borrower, and how much to keep. Money that is kept is retained safely for the Investor. Whatever money is passed from the Investor is trebled before passing to the Borrower. At that point the Borrower must decide how much to return to the Investor. The maximum total result arises when the Investor invests all of the allocation with the Borrower. The Investor determines the total value in a round and the Borrower decides how to allocate the proceeds.

In addition to the separation of moves of the Investor and Borrower, there is a separation of role. The Investor decides how much to trust the Borrower, and the Borrower's trustworthiness is tested. This game is not usually played as a single round game; a number of rounds are played, over which the Investor and Borrower continue to benefit from their respective investments of trust and trustworthiness.

The Investor Games reduces the degree of confounding of trust and cooperation. In the Prisoner's Dilemma both players are equally exposed to each other and both players move together. This is not common in negotiating trust, which tend to be sequential, trust decisions and subsequent activity tends to be asymmetric - one person decides to trust and acts on the decision, and the second person then may choose to respect that trust or take advantage of it. Unlike the PD, the Investor Game is asymmetric and asynchronous.

These more recent experiments with trust games have been used to explore variation in the operation of trust across cultures (Yamagishi et al., 1998; Yamagishi, 2003), the importance of other factors such as norm enforcement (Yamagishi et al., 1998), the use made of available information (Yamagishi, 2001) and the role of perceived fairness (Camerer, 2003a), including fairness in the size of investment and returns expected (Rieskamp, 2001).

While experiments using trust games involve more trust than does the Prisoner's Dilemma, they do not isolate trust. In trust games, the first to move is given some money, and can decide to keep it all, or to pass some or all to the second to move. The second to move receives the investment plus a premium, then decides how much to return to the first player. A decision by the first player to pay money to another may be a result of a trust decision but, equally, it might be due to altruism, or a belief in equal sharing (Cox, 2004) or some combination of these and altruism. In another game, called the Dictator Game, the first player is given some money, and decides how much to pass to the second player. The second player returns nothing, so the game is devoid of any component of trust or reciprocation. Using a comparison of player behaviour in the Investor Game and the Dictator Game, Cox (2004) has identified that the trust game does produce different behaviour to the Dictator Game.

Generalised trust: Traits and norms

While the previous sections have been concerned with identifying the cognitive mechanisms involved in specific instances of trust, they did not discuss individual differences in trust. Somewhat paradoxically, individual differences in trust tend to be revealed in a construct that is often labelled generalised trust. Repeating the definition from the first section of this chapter, generalised trust is a non-specific expectation about the behaviour of others, and so might be thought of a belief about norms of trustworthy behaviour.

Trust as a disposition

The inability to differentiate trust and cooperation in early game experiments led to a move away from experimental methods in investigating trust (Cook & Cooper, 2003; Goto, 1996). Rotter (1967) developed an Interpersonal Trust Scale (ITS) and he and others used this and similar scales to investigate the correlation of trust with a variety of personality and social constructs. This research has produced a body of survey results in which interpersonal trust is seen as a dispositional or personality variable.

While both experimental and dispositional streams of psychological research into trust seem compatible with everyday ideas of trust, they did not produce a coherent picture of trust. Studies incorporating both experimental and ITS measures have suggested that trust as represented by the PD and the Interpersonal Trust Scale were not correlated, indicating that these accessed significantly different concepts (MacDonald, Kessel, & Fuller, 1972). This is not entirely surprising, given that the PD includes features of both cooperation and trust. Similar concerns exist with the Interpersonal Trust Scale. This asks about trust in a number of specific institutions, alongside more general elements of trust. It may be that these streams cannot be merged, because both package trust alongside other constructs. It has also been suggested that the findings from these research streams cannot be merged because they miss essential elements of trust. For example, sociologists might take the view that trust is an irreducible quality of social groupings (Lewis & Weigert, 1985), and the complexity of different forms of trust and their embeddedness in social relationships might prevent the merging of the various streams of trust research.

This may be unduly pessimistic. The content of the ITS questions are particularly concerned with beliefs about trust, rather than about the practice of trust. Beliefs are particularly associated with rational thinking (Evans, 2003), and this tends to position generalised trust as a component that may be drawn on when we use System 2. If the practice of trust usually draws on a System 1 assessment and heuristic type decision, it may be that it bypasses beliefs. If System 2 does process differently to System 1, and has been found valuable for its ability to check and override System 1, then we should not expect to find System 1 practice necessarily correlated with System 2 beliefs.

Generalised trust

In economics research generalised trust is discussed in terms of a propensity to trust strangers, and people that we do not know well. This does not suggest any discrimination in the application of trust, and so does not differentiate between trust and gullibility. In this form, the concept of generalised trust lacks credibility as, in practice, people are highly unlikely to risk a lot on a completely unknown stranger.

Nevertheless, people do give positive answers to questions like the generalised trust question from the United States' General Social Survey ("GSS 2000 Codebook.", 2000). This question asks "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" Similar questions have been used in international surveys of social opinion such as the World Values Survey (Inglehart, 1997). In these surveys, this question has been interpreted as accessing generalised trust. Although many people agree with the statement that most people might be trusted, most answering this way would not apply this universally and indiscriminately. Hardin (2002) claims that generalised trust essentially does not make sense. He maintains that we are selective about who we trust and with what, and so questions that ask about trust that do not specify these parameters do not make sense. Notwithstanding these comments, he does discuss individuals having general expectation of the trustworthiness of other people, where trustworthiness is the extent to which the person being trusted will behave as expected(Hardin, 2002). This expectation would appear to have much in common with generalised trust as accessed by the GSS question. Generalised trust might better be described as a generalised expectation of the trustworthiness of others. The questions in the ITS and the survey questions on generalised trust are so similar as to suggest that they might be accessing the same construct. Thus the generalised trust question is accessing peoples beliefs about norms of trust and trustworthiness.

The GSS question on trust has also been criticised for conflating two constructs; the question offers trust and caution (Miller & Mitamura, 2003) as an either/or bipole. Respondents are allowed two possible responses to the generalised trust question: "Most people can be trusted" and "[you] can't be too careful." These two responses are differently framed as trusted and careful, and they may tap two separate constructs. Rather than answering a question about trust, it may be that people are responding to the word caution rather than reading the trust and caution clauses as antonym alternatives.

A further criticism is that many studies accessing generalised trust either use a single explicit question about whether, in general, people can be trusted or, alternately, questions about generalised trust are bracketed with questions on trust in the government and its institutions and with questions on optimism. This grouping of questions closely parallels the ITS; both produce a multidimensional measure of trust, including trust of unspecified people alongside trust of institutions like the government or the clergy. More recent psychological measures of trust as a disposition have been more uni-dimensional (Couch & Jones, 1997; Omodei & McLennan, 2000) to avoid this problem, but there has been no equivalent move in the design of questions on generalised trust.

Although the design of the generalised trust question raises questions about what construct is being tapped, it would seem that people are prepared to answer the question. And, in the aggregate, their answers seem to have some implications for the welfare of their countries. At the national level, the tendency to agree that people can be trusted correlates with other important national social, political and economic indicators. The mechanism by which this may happen remains contentious. Some analysis has been done assuming that the cost of a lack of trust is a simple sum of the costs of individual transactions (Zak & Knack, 2001), although these authors do note that direct economic costs may be one of the lesser consequences of a lack of trust in a society. Others suggest that trust only has meaning at a societal level (Lewis & Weigert, 1985), and so cannot be accessed as an aggregate of individual decisions.

Learning trust

Despite the heading of this section, it is not at all clear that we can learn to trust. Babies and young children are entirely dependent on their parents at birth and for many years, and they must be able to trust to survive. This being the case, it is likely that our initial state is one of unquestioning trust (Baier, 1986). There is not much doubt that we can learn distrust from subsequent experience. Erikson identifies learning to trust and distrust as the first major developmental task, but learning to distrust is a continually refined: as long as we continue to trust, and continue to provide opportunities to test that trust we may find situations in which it is unjustified.

This learning process is not symmetrical. It is less obvious that we can learn to trust by trial and error. If we do not trust, we do not have an opportunity to test whether our trust would have been honoured. If we distrust we cannot learn trust through experience. We may have, nevertheless, retain opportunities to learn trust through observational learning.

Another, less obvious, possibility derives from the analysis that suggests that trust might best be thought of as an attitude. Attitudes can be changed by changing our behaviour, so maybe it is possible that trust can develop through behaving as though we trust, even though we may not hold this attitude. Baier (1986) considers the effect of this situation on the trustee, who may respond with increased trustworthiness, but does not extend this to considering the possible effects on the truster.

Summary

This chapter has described a number of directions from which trust has been explored and described. The variety of ideas surrounding trust, and of approaches taken to researching trust, highlights that trust is a rich concept, having social, emotional, and cognitive dimensions. A key element is that, when faced with incomplete knowledge about another's intentions, there is a belief that those intentions are benign, and action in accordance with that belief. A further key element is that there is a degree of risk if these beliefs are wrong. While these are consistent themes in defining trust, particularly those concerning risk and a lack of knowledge, the precise meaning of trust varies across particular situations. For example, trust in receiving support from the family is quite different to trusting an anonymous postal worker with a package.

An ability to function under such conditions is useful in enabling cooperative action which has the potential to generate greater overall benefits for the cooperating parties. This applies particularly in situations in which the person is not well-known to us. In this case we cannot make well-informed predictions about that person's likely behaviour based on our past experience with that person. If we are only prepared to deal with people whom we can predict will behave honestly based on personal knowledge, we are restricted to cooperating with people whom we know very well, such as our immediate family. Trust allows humans to benefit from cooperating with others outside their immediate family. But, for trust to work the individuals that a person is likely to encounter must share the same understandings of trust, trustworthiness, and social norms. While individually founded, cooperative action needs shared understandings and norms around trust to also develop within networks.

Placing trust into an agent-based model

Modelling trust using theory that is specific to trust, I was faced with an immediate challenge. As noted earlier, there are a number of possible constructs that we might model. Translated to a modelling perspective, the possibilities include incorporating trust into agents as a personality factor, an attitude, a decision-making task, or as a behaviour. Even with a particular construct in mind, there is relatively little theory that is explicitly about trust. There is even less in such a form that it can be adapted to a trust simulation task. Trust as a personality factor provides a good example of this. While a number of scales that measure trust as a personality factor have been proposed, they do not necessarily predict individual behaviour in specific situations. For example, behaviour in the Prisoner's Dilemma game is not correlated with measures of generalised trust (MacDonald et al., 1972). For the modeller, this means that if, for example, we have a component that specifies the degree of a personality factor *generalised trust* associated with an individual agent, there is little to suggest how this influences the individual agent's interaction with other agents and its environment.

Trust has such an important role in allowing people to act cooperatively that it is pervasive feature in social interactions. It sometimes seems to me that the extent of this pervasiveness is such that people do not necessarily notice that trust is operating. For example, people who would claim that they do not trust anyone don't stop to consider paying for their burger before they have received it. Possibly another sign of its pervasiveness is that people tend to have a shared understanding of what trust means in practice.

Despite shared popular understandings about what trust means, there are no universally accepted formal academic definitions of what trust is. This is because a diverse range of conditions and elements are required for trust to be engaged, and recognised as being trust. As a result of this diversity, rather than a single academic definition, there are a collection of ideas about what trust is, each of which makes some assumptions about the situation to which it applies.

The diversity in trust as a concept extends to the forms in which trust manifests. This chapter has discussed some of these forms. Any of these various forms might be suitable for inclusion in a simulation model, although some, such as a disposition to trust as a personality factor, might be more applicable to incorporating trust as a static characteristic within a model that addresses something other than trust. An example might be the models of the economic costs of varying degrees of prudential activity that are triggered in part as a result of the amount of generalised trust in individuals in a population.

The foregoing has identified some of the ways that trust might occur in individuals, that might also be carried into representations within agents. These are specific to trust. A model of trust in networks might use any one of these in constructing agents. For example, the agents might be equipped with a algorithm for making trust decisions that is derived from research into the cognitive mechanisms of trust.

A second approach might be to identify more general ideas from psychology that might be incorporated into individual agents. There are a range of possible candidate theoretical ideas. For example, theories of learning might be applied to the learning of which agents can and cannot be trusted.

The next two chapters describe the development of two models of trust. The first is based on a game theoretical finding about strategies that are engaged in the Investor Game, extending this model to a population all of whom use these strategies. The second is based on a naturalistic model in which agents are located in an online auction setting, provided with the same information that is provided in real online markets, and with a means of exchanging strategy information.

Chapter 7

The Basic Breaking Model

The first of the two agent-based models of trust that I constructed was based on a formal game that involves trust, called the Investor Game. Strategy heuristics have been identified for this game, using a combination of experiment and simulation (Rieskamp, 2001).

Formal games

The simplicity of formal games, and the ability to introduce information in a controlled way makes these useful devices for experiments in social psychology, as the games can be modified to introduce other elements in a controlled way for experiments. Game experiments can produce theory that is suitable for direct application to a simulation model, in the form of strategies, and factors that have been found to modify these strategies. One of only a few specific models of trust as individual action has emerged from research using formal trust games.

One attraction of using games is that many games have been analysed formally. The task set participants is well understood, and in many cases theoretically optimal solutions have been identified. These solutions can provide a cognitive baseline: if players were to play the game using entirely rational thinking, and thinking only within the immediate context of the game, they should use optimal strategies.

Game theory was initially developed as a class of mathematical problems by John von Neumann (Camerer, 2003a, p.2) revolving around the strategic interaction of two players. Von Neumann recognised these games as having application in strategic decision-making. In game theory, the games are stripped down and formalised interactions, with small rule sets and very restrictive rules. These formal games can be applied to a variety of real world situations involving strategy and negotiation. Examples include plea bargaining, auction design, the negotiation of pay scales, and the distribution of the ownership, costs, and benefits of common goods.

Economics has made wide use of game theory. In this discipline, game theory is a natural fit with the subject, both because of its numerical formulation, and because of economics interest in the production and distribution of scarce resources. But game theory has also been applied in other disciplines in which strategic interactions arise, including psychology.

As noted above, optimal solutions can be identified for many games, but in experiments using formal games, a substantial proportion of people do not play as if they are using optimal strategies. On the assumption that it is rational to choose the moves that give an optimal outcome, many players appear to use irrational strategies. One example is provided by the Dictator Game. In this game, a player is given an amount of money, and told that he may give as much, or as little, as he chooses to the other participant. The optimal solution is for the player to offer nothing, thus maximising what he keeps. But, rather than keep it all, many human players will give the other player at least some of the money (Camerer, 2003a, p.57-58).

The Ultimatum Game extends the Dictator Game, this time giving the second player an opportunity to reject an offer that she finds unsatisfactory. The second player can accept the offered amount, or reject it, in which case both players gain nothing. The optimal response is to take whatever is offered, as this is better than nothing. Despite this, many players will reject offers that are too low. In the Ultimatum Game many first move players offer an amount in the 40-50% range (Camerer, 2003a, p. 50-55). This suggests not only that players are likely to play in an apparently irrational way, but that this is expected by the player making the offer.

Both games demonstrate situations in which individuals make moves that are not optimal, at least in terms of the objects of the game. There are a number of possible reasons that people may make these, apparently irrational, choices, including social factors. In the Ultimatum Game example, for example, players may be playing in accordance with norms for sharing. Further, it appears that in real world settings people can access the gains that can accrue from playing cooperatively, although these strategies may not be optimal in a formal mathematical sense. This demonstrates an obvious problem with formal game theoretic solutions, particularly in economic modelling: These tend to assume economically rational behaviour. There is little evidence to suggest that this assumption is reasonable. Experiments using formal games have served both to demonstrate that people do not approach games using purely rational thinking, and as one means to explore which of other possible approaches people might be using.

Games can be played either as a one off interaction, or as a series of repeated rounds. Repeating the game can result in a drastic change in the strategies used by the players. This is particularly pronounced for trust games, in which there is a potential benefit from cooperation, and so a payoff for demonstrating trustworthiness is possible.

The Investor Game

The Investor Game is a game that engages trust for one of the players. In this game (Camerer, 2003a, p. 85), the two players are given \$10 each. The first player, the investor, decides how much to invest with the second player, the borrower. This may be any amount, from nothing to the full \$10. The borrower receives three times what the investor decides to invest, plus \$10: an amount between \$10 and \$40. The borrower then decides how much to return to the investor. Again, this may be any amount between nothing and the full \$40. Even if the players are restricted to whole dollar amounts, there are a large number of possible outcomes - either investor or borrower might finish with any amount between \$0 and \$40.

The best overall outcome depends on the investor investing the whole \$10. But to achieve this best overall outcome, the investor must have an expectation that the borrower will return at least the \$10 invested. This involves trust on the part of the investor, as he gives the borrower the full amount possible, in the face of a risk that the borrower may not not return this. Proponents of the Investor Game as a trust game claim that use of this game in trust research can isolate trust by removing other cues that may affect a decision (Camerer, 2003a, p. 85), such as looks, gender, and how the person communicates. This can make formal games useful as the basis for trust experiments, as they can isolate a minimal version of trust, the trust decision, to which other factors may be added. Social psychologists have used this ability to test other factors that can influence such a decision.

Hardin (2002) firmly places trust as a cognitive process, rather than the enablement of a behaviour. More specifically, he identifies trust as an unconscious, as opposed to a conscious, decision process (Hardin, 2002). According-ing to the dual process model (S. A. Sloman, 2002), unconscious decision-making is characteristic of a heuristic, rather than a rational decision-making process. This suggests both that we do not deliberate on a decision as to whether or not to trust. That decision is likely to be heuristic, rather than a rational decision solving some sort of rational evaluation algorithm. Unlike the separation of cognitive process and consequent behaviour, the distinction between conscious and unconscious process may be less significant for the model, but more for the form of the decision-making process. The formulation of trust as a cognitive process is an appropriate choice for a theoretical basis for an agent-based model.

This model is a representation of a set of heuristics that were identified as optimal strategies for the Investor Game by Rieskamp (2001). These were derived from the interactions observed in an Investor Game experiment with human players. Rieskamp (2001) inferred individual strategy heuristics through comparing experimental results with a set of strategies derived through an agent-based model of the Investor Game. The interactions of people in an Investor Game experiment were clustered, producing a set of patterns of interaction among human participants. An agent-based model was used to identify a set of optimal strategies, and the patterns of interaction generated by agents using these strategies. These two sets of patterns were compared, identifying patterns generated by humans with strategies developed by simulation agents.

Each agent in the model is allocated one of two optimal investor strategies, and one of two optimal borrower strategies. Of these, only the investor strategies depend on trust (Camerer, 2003a). Each round investors are given an amount of money. Investors decide how much of this to invest with the borrower. For each dollar invested, the borrower receives three times the amount. The borrower then decides how much to return to the investor. For example, if investors are given $1 each round, and the investor decides to invest all of this, the borrower then receives $3. The borrower then decides how much to return. Any more than $1 returned represents no loss to the investor, but an inequitable share of the profits. All decision thresholds are drawn from (2001).

The theoretical profit-maximising solution for one-off interactions is that the borrower, if it does get given any money, acts in an untrustworthy manner, keeping the entire of what it is given. This being the expected rational strategy for borrowers, the investor does not trust the borrower at all, and so keeps all of the money. There is no potential profit for the borrower in demonstrating trustworthiness, as there are no future interactions between the players in which to collect the benefits of cooperation.

While entirely rational, this pair of strategies are far from optimal. Both players forgo the gains that can be made through cooperating. As described above, the game produces \$1 in total per round. If the investor does invest the total available money, a round produces \$3 in total, split somehow between investor and borrower.

These rational strategies do not hold for repeated (iterated) games, in which players do not know how many times they will partner the same player. In these games, there is future value for the borrower in demonstrating trustworthiness. What the borrower forgoes by returning money in a game should be more than covered in subsequent rounds.

The four strategies can be described in the state diagrams Figures 7.1-7.4. The investor strategies are:

1. Hesitant (Figure 7.1). The investor begins cautiously, investing 50% of its income for the round (state 1). Whatever the outcome of this first investment, the investor again invests a second time (state 2). If more than 33% is returned, the investor assesses that the investment has been reciprocated, the agent remains in state 2. If it is not, the investor does not invest in the next round (state 3), after which it returns to state 1.

Investor Game model

1. Moderately-Grim (Figure 7.2). The investor begins by investing all of its income for the round (state 1). If more than 33% is returned by the borrower, the investor assesses that the borrower has reciprocated. If less, the investor assesses that it has been exploited. If the investment was exploited, the investor refuses to invest any more (state 3). If the investment was reciprocated the investor again invests 100% (state 2). From this point, if the borrower defaults, the investor reverts to state 1, otherwise it remains in state 2.

The borrower strategies are:

1. Reactive (Figure 7.3). The borrower assesses whether it has been trusted round by round. If the borrower receives more than 17% of

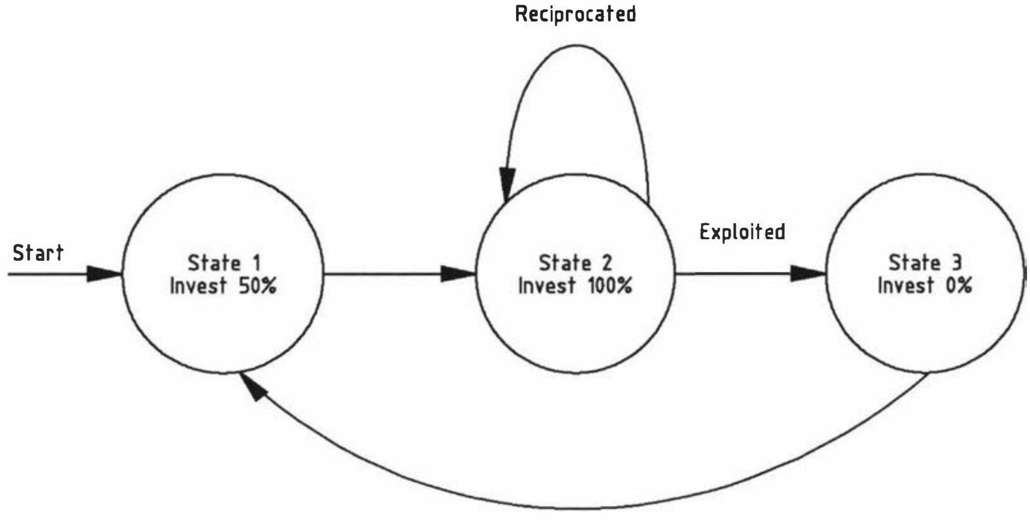

Figure 7.1: Hesitant Strategy State Diagram (Rieskamp, 2001)

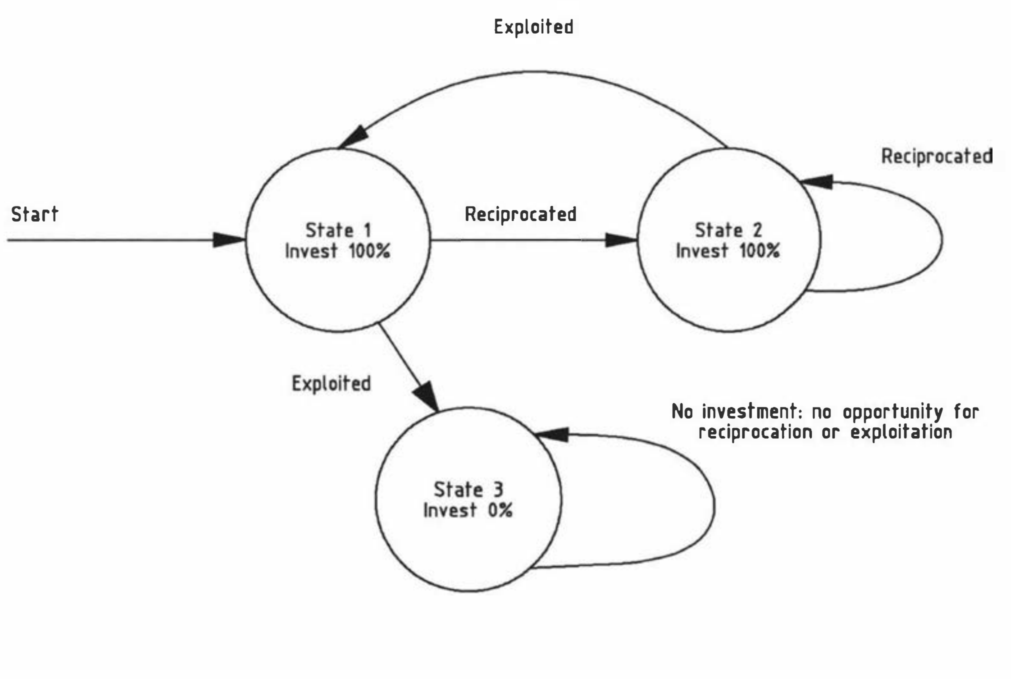

Figure 7.2: Moderately-Grim Strategy State Diagram (Rieskamp, 2001)

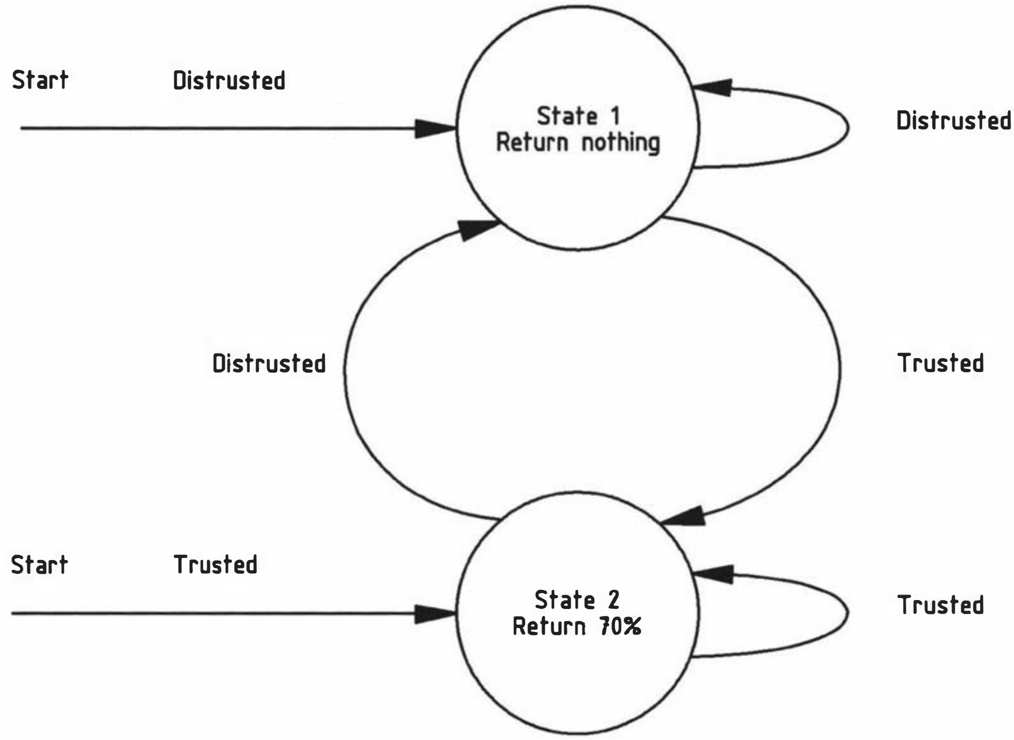

Figure 7.3: Reactive Strategy State Diagram (Rieskamp, 2001)

the investor's income, it returns 70% of the total amount received. If the borrower receives less than this, it assesses this as indicating mistrust and returns nothing.

1. Half-Back (Figure 7.4). The first round the borrower returns half of the amount received (state 1). While it continues to receive more than 12% of the investor's income, it remains in this state. If in subsequent rounds it receives less investment, it returns nothing until it is again trusted (state 2). If it is subsequently trusted again, it returns to state 1, and returns the money.

Camerer (2003a) notes that it is much simpler than the situations in which real world trust decisions are made, but claims that this is because the model is a pure form of trust decision. It describes a simple set of rules for making decisions, and provides no other sources of information for agents to draw from, such as appearance, or any interaction beyond passing money back and forth.

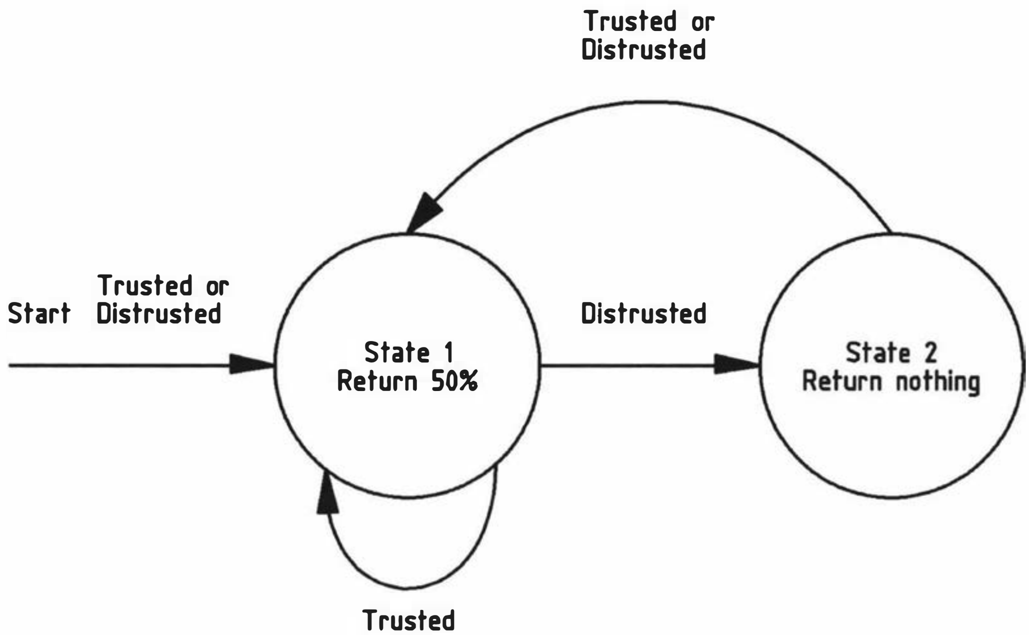

Figure 7.4: Half-Back Strategy State Diagram (Rieskamp, 2001)

It is possible that unexpected patterns might flow on from agents trading from this simple set of rules. Patterns arising from a large number of simple trust interactions is one of the mechanisms by which the richer patterns of trust might be generated. If such patterns were emergent, almost by definition, we would not be able to predict them until they were observed. This model is therefore a pure exploration as to whether patterns do develop from a large number of agents carrying out simple interactions that rely on trust for success.

Building the Investor Game model

Working with an agent-based model tends to be an iterative process. A section of programme code is written, then run. This acts both to test that the programme is doing what we expect it to, and to add elements to the model in a step by step manner. Later in the modelling process, this iterative process becomes dominated by experimentation.

The first task was to populate the model, provide the agents with the basic set of four strategy heuristics, and test that these were working as intended. A quarter of the total agent population is allocated to each of the four strategies. Investors pick up a borrower partner and vice versa. In the initial version, each agent remains with the same partner until the end. In the final version, they can break unsatisfactory partnerships and remake partnerships with other unpartnered agents.

Each of these strategies has parameters marking the boundaries between trust and mistrust, between reciprocity and exploitation, and fixing the levels of investment and return that were drawn from Rieskamp (2001). These were used for the first version of the model. In later versions, variability was added to these, so that there were individual expressions of the strategies.

The description of the final version of the model that follows includes both the individual variation in strategy parameters and the ability for agents to break off trading where the partnership is unsatisfactory. The description is divided into two major programme sections: one representing the abilities and characteristics of the agents, and one that both represents the environment and manages tasks common to all agents. These include managing turn taking and communication between the agents, and handling data output. In the listing of the final version of this model provided in the appendix, the agent section of the programme is called TrustAgent, and the model section is called BasicBreaking.

An Investor Game model: Basic Breaking

The Basic Breaking model has four versions, each adding one feature to the previous version. There are a number of reasons to work this way. Firstly it is helpful in developing and software verification of the model and programme. In the initial stages, a minimal model allows the modeller to be able to track how the programme works step by step. The programme is run with a small number of agents, and output can be gathered after each step to ensure that the programme and model is acting as expected at each step. The understanding from each step acts as a foundation for following steps.

The first Basic Breaking model, BasicHeuristic, provides the agents with the strategies exactly as given by Rieskamp (2001). In this model, the agents with the same strategy exercise this in exactly the same way. That is, the levels at which investor agents decide that an investment has been reciprocated, and borrower agents decide that the investor has trusted them are fixed for all agents.

BasicRandomised adds some variability, by replacing the fixed threshold levels at which agents judge reciprocation and trust to have occurred with randomly determined threshold levels, with a mean equal to the level used in the BasicHeuristic model.

Moderately-Grim investors in the BasicHeuristic model can become locked in a cycle of distrust. If the borrower either fails to return enough of the first investment, or fails to return sufficient for two cycles in a row, the Moderately-Grim investor will not invest any more. Either of these results in the Moderately-Grim investor refusing to trust this partner again. This version allows Moderately-Grim investors to break partnerships in which the borrower does not return enough of the investment.

While BasicBreaking4 introduced this breaking component, when the model was run no breaks were generated. The actions of the agents are such that investor agents always provide an initial opening for cooperation, and borrowers always reciprocate. The BasicBreaking4 model allowed only the overall number of agents and the proportion using each of the four strategies to be varied. The BasicBreaking5 model allows the decision set-points within the strategies to be varied.

TrustAgent

This section of the model defines the agent which, in this model, represents an individual investor or borrower. Each agent attempts to accumulate wealth each round through a successful investment and repayment transaction. Each agent has an identifying number and an allocated role (as investor or borrower) and corresponding strategy. Each records the total wealth that it has accumulated.

Each agent also has a set of characteristics setting parameters in its decision-making heuristic. They set the levels at which the agent regards its partner as having trusted it (for borrower agents) and at which the agent regards its trust having been justified through reciprocity having been satisfied (for investor agents). In the first version of the model, this was fixed for each agent at the level reported by Rieskamp (2001). In later versions, variation was added to allow for agents to have individual variation in these.

Table 7.1: *Gains by investors and borrowers at each round, borrower thresholds set at 0.1 and 0.4.*

		investor	borrower
hesitant	reactive	21	9
hesitant	half-back	15	15
moderately-grim	reactive	21	9
moderately-grim	half-back	15	15

Results

The results were analysed using the R statistics (Ihaka & Gentleman, 1996) package. Three analyses were carried out. The first investigated the effects of varying the thresholds at which agents determine whether they were trusted or whether the partner reciprocated. For this analysis, every agent had the same threshold settings. The next analysis randomly allocated thresholds for each agent.

Varying the thresholds, with no randomisation

The first model used a population of 100 agents. The model was run as a batch to generate a combination of each of 4 different levels for each of 3 different threshold values (reciprocity for investors, and trustedness for Reactive and Half-Back borrowers). This produced a total of 64 runs, each of which was allowed to run for 10 rounds. For each agent, the reciprocity or trustedness decision thresholds were set to 0.1, 0.4, 0.7, and 1.0. The amounts invested and returned were kept constant, and equal to the figures in (Rieskamp, 2001).

Results

There are four strategy combinations: Moderately-Grim and Reactive; Moderately-Grim and Half-Back; Hesitant and Reactive; Hesitant and Half-Back. For each of these combinations, investor and borrower earnings are fixed at each cycle after the first. While thresholds for determining trust are low, the total earnings are three times the investment made by the investor, and the borrower's strategy that determines the distribution of the income, as Reactive borrowers return 70% of what they receive, and Half-Back borrowers return 50% of what they receive.

Table 7.2: *Gains by investors and borrowers at each round, borrower thresholds set at 0.7 and 1.0.*

		investor	borrower
hesitant	reactive	0	9
hesitant	half-back	0	15
moderately-grim	reactive	21	9
moderately-grim	half-back	15	15

Discussion

As we might expect from a model in which there is no scope for variation, and no scope for dynamic responses, this model produced constant return to investor and borrower at each interval. The first interval differed for some agent pairings, as the Hesitant strategy begins with a lower investment at the first step than the Moderately-Grim strategy. Where the TrustThreshold is higher than 0.5, Hesitant investors never meet the trust threshold of borrowers, so pairings with Hesitant investors produce worse outcomes

Individual variation in strategy realisation

The second analysis was of the outcomes in a population of 100 agents after the 100th round. The decision thresholds, and the amounts invested and returned, were randomly distributed around the mean figures reported by (Rieskamp, 2001). These were applied to a population of 100 agents. The BasicBreaking5 model was first used to produce an image of the outcomes when the agents use strategies based very closely on the strategies extracted from the paper.

Result

The borrower's strategy has a significant main effect on the gains made by borrowers, $F(1, 45) = 17.77$, $p < .001$, such that Half-Back borrowers accumulate more, $M = 1386.4$, $SD = 423.9$, than Reactive borrowers, $M = 879.7$, $SD = 397.0$. There was no significant main effect of the investor partner's strategy on borrower gain, $F(1, 45) = 0.119$, $p = 0.732$, ns, between with Moderately Grim, $M = 1148.1$, $SD = 503.8$, and Hesitant partners, $M = 1106.8$, $SD = 463.7$. There is no significant interaction effect, $F(1, 45) = 0.007$, $p = .932$, ns.

There is also a significant main effect on the investors income depending on the strategy used by the borrower, $F(1, 45) = 14.77$, $p < .001$, such that investors with Reactive partners accumulate more, $M = 2103.4$, $SD = 655.6$, than investors with Half-Back partners, $M = 1472.1$, $SD = 439.9$. Again, there was no significant main effect of investor strategy on gains made by investors, $F(1, 45) = 2.54$, $p = .12$, ns, between Moderately Grim investors, $M = 1669.0$, $SD = 632.4$, and Hesitant investors, $M = 1924.6$, $SD = 634.1$. There is again no significant interaction effect, $F(1, 45) = 0.003$, $p = .96$, ns.

Discussion

In the strategy described by (Rieskamp, 2001), Reactive borrowers who have been trusted return a mean 0.7 of the amount received, and Half-Back borrowers who have been trusted return a mean of 0.5 of the amount received. This corresponds to a ratio of 1.4: 1 in the returns from Reactive and Half-Back partners respectively. Results from the model suggests that the ratio of mean returns for investors was 1.58 : 1.

Despite this model providing for unsatisfactory relationships to be broken off, the strategy does not generate any relationship breaks. All investors attempt some investment at the first step, and all borrowers reciprocate these investments.

Sensitivity to set-points

As noted previously, there are a number of constants noted in the strategies identified by (Rieskamp, 2001). One group sets the size of the original investment, and the amounts to be returned. The other group sets the levels at which borrowers accept that they have been trusted. The original figures were 12% of the investor's income (for Half-Back agents) and 17% of the investor's income (for Reactive agents). Investors set the point at which they accept that their trust has been reciprocated at 34% of the income received by the borrower. Once the investor's investment is tripled, the original investment corresponds to a third of the amount received by the borrower. The investor requires that this investment be returned before accepting that the borrower has reciprocated.

Within the broad logic of the strategies modelled, there are few ways that failures of trust and reciprocity can be generated. The amounts invested and returned can be varied, and the set points at which trust and reciprocity are determined can be measured. These are alternate ways of triggering the same effects, for example if the initial amount invested is dropped, eventually it will fall below the threshold for having been trusted, and the same effect occurs if the threshold is raised. These being equivalent, the sensitivity to these constants was explored by varying just one set of constants, the set points.

This model models a series of interactions in a population of 100 agents. A 10 step run is carried out for each combination of four levels (0.1, 0.4, 0.7, 1.0), of the three threshold settings (reciprocity for investors, trust threshold for Reactive borrowers, trust threshold for Half-back borrowers), resulting in 64 runs of 10 steps each.

Results

Two standard step-wise linear regressions were carried out, one with the borrowers accumulated gain as dependent variable, and the other with the investor's accumulated gain as the dependent variable. The independent variables applied to both regression models were the strategy of the borrower, the strategy of the investor, the threshold at which reciprocity is determined, the thresholds at which Reactive agents determine that they were trusted, and the thresholds at which Half-Back agents determine that they are trusted.

moneyBorrower All variables except the investor's strategy generate significant estimated coefficients in the regression on the borrower's money. Comparison of the shift in the adjusted R^2 score as variables were removed showed that removing all variables but three had little effect on the overall model. The three remaining variables are strategyBorrower, TrustThresholdHalfBack, and TrustThresholdReactive, corresponding to the borrower's strategy, and the thresholds at which borrowers determine trust. Table 7.3 lists the estimated regression coefficients for the remaining variables.

The overall model has an Adjusted R^2 of 0.193. Although each of the independent variables is significant, the overall model explains relatively little of the variance in borrowers' incomes.

moneyInvestor Again all variables have significant estimated coefficients when regressed on the gains made by the investor. With all variables in-

Table 7.3: *Linear regression coefficients, variables influencing borrower gains.*

	Estimate	Std. Error	t value	p
(Intercept)	41.819	3.414	12.249	< .001 ***
strategyBorrower	41.940	1.749	23.973	< .001 ***
TrustThresholdHalfBack	19.779	2.607	7.587	< .001 ***
TrustThresholdReactive	28.223	2.609	10.818	28.223

Significance codes: *** < 0.001

Table 7.4: *Linear regression coefficients, variables influencing investor gains*

	Estimate	Std. Error	t value	p
(Intercept)	181.981	6.825	26.664	< .001 ***
ReciprocityThreshold	-36.221	3.898	-9.293	< .001 ***
strategyBorrower	-38.332	2.615	-14.657	< .001 ***
strategyInvestor	62.522	2.615	23.908	< .001 ***
TrustThresholdHalfBack	-51.861	3.897	-13.309	< .001 ***
TrustThresholdReactive	-78.270	3.900	-20.071	< .001 ***

Significance codes: *** < 0.001

cluded in the analysis, the adjusted R^2 is 0.32. Removing variables from the regression equation steadily reduces the adjusted R^2, indicating that the best overall model incorporates all of the variables. Again, while all coefficients are highly significant, the model accounts for only a moderate proportion of the variance in investor income. Table 7.4 lists the estimated regression coefficients for the remaining variables.

Discussion

The positive sign on the estimated coefficient for strategyBorrower (Table 7.3) indicates that borrower gain depends on borrower strategy; on average Half-back borrowers made greater gains than did Reactive borrowers. This is consistent with the previous ANOVA analysis. The positive sign on the estimated coefficients for the borrower thresholds indicates that borrower gains increase as borrower thresholds increase. As the borrower thresholds increase, the chances of trust being recognised drop, and thus the investor may not get any money returned.

The gains made by investors depend on the outcomes of three stages: the amount of the initial investment, the amount returned by the borrower, and the subsequent assessment made by the investor that affects the next transaction round. The regression result for investor income is consistent with this. The three variables of the borrower income are again significant variables in setting the investor income, but with the opposite sign. This reflects that an increase borrower incomes tends to come at the cost of investor incomes, as the borrower income is the difference between the total amount received and the investor income.

The relatively low proportion of variance in borrowers' gains explained by the model is noteworthy. This may be generated, not by a change in the process, which is controlled by the four strategies, but by the individual variation in the application of these algorithms. The individual constants were scattered around the mean, using a normal distribution with a standard deviation set at about 20% of the mean. This illustrates that individual variation is a source of noise in its own right.

The regression shows that for a game of this structure, investors gains depend on both the borrowers strategy and their own. It also depends on the levels at which agents determine that trust has been accorded and reciprocated. For investors, cooperation depends both on individual strategies and on the assessments that both investor and borrower make about whether trust has been accorded and reciprocated.

This is not symmetric for borrowers, for whom the greatest influences are their own strategies and their own assessments as to whether they have been trusted.

Chapter 8

The Trading Model

In looking for theory to apply to an agent-based model of trust, I found an almost paradoxical dichotomy: there are an embarrassment of riches in the wider research findings of social psychology, many of which are potentially applicable to trust, but few specific models of trust. The previous chapter described the development of a model that was based on a cognitive model that is specific to trust in an abstract setting.

Empirical research has supported a vast array of theories that in a specific social situation people will, on average, make particular assessments of others, and behave in particular ways. Alongside social psychology, cognitive psychology offers theories about how people gather information and make decisions. This chapter describes the development and results from a model that is based on more general ideas from social psychology, applied to a naturalistic setting.

Whether calling on a specific theory about how trust functions, or a more general psychological theory, I encountered a common difficulty in incorporating psychological research into an agent-based model. Despite the vast body of psychological knowledge and findings, much of this theory is not in a form that makes it suitable for direct incorporation into agent-based models of trust. There are relatively few theoretical ideas, in the form that they exist in psychology, that can be applied directly in developing well-supported individual agents. For example, a finding that wealthier people are, on average, more trusting (Hout, 2003) might be reported in terms of means and standard deviations, and the results of hypothesis tests about these means and standard deviations. In some circumstances, applying the mean value to all members of a population might be a reasonable representation. At other times, the modeller is likely to want to represent diversity in the population in a credible way. While analyses have often been carried out under the assumption that the distribution was normal, the shape of the distribution or even the range of values found, is often not reported. Even when using psychological theory that is well supported, without distribution information, the modeller is unlikely to have sufficient information to be able to incorporate theory directly into a model. This means that the modeller will, almost certainly, have to use judgement, and make assumptions at some point in the modelling process.

Trust in online auctions

The model described in this chapter is based around trades in an online auction market, such as eBay or New Zealand's Trade Me. The sale process in these online markets begins with a seller offering an item for sale. Prospective buyers must decide whether or not to place a bid. This decision can be broken down into two decisions. The first decision corresponds to a potential buyer deciding whether or not they are interested in buying the item. In the trading model, this is modelled as an entirely random auction process, in which the seller randomly chooses a selling price, and each bidder decides how much they are prepared to bid for it.

If a trader is interested in buying the item, the trader then has also to decide whether or not they are prepared to trust the seller. The purchase process usually involves the buyer paying the seller, and the seller makes delivery once the payment has been received. It is possible that, having received payment, the seller will not behave in accordance with the buyer's expectations, and may not deliver the item. Therefore, there is a component of risk for the buyer, who may pay the money to a dishonest seller who may not deliver the item. Potential buyers are typically provided very little information on the seller. The only information that is available is the history of previous trading behaviour, the numbers of good and bad trades, and the comments made by other traders. There are none of the everyday social cues that people might use in face to face trading.

There is a second source of risk for both buyer and seller, beyond dishonesty. Trades can fail if unreliable traders do not communicate, or if they do not make payment for an item that they have bid on. In this case, the direct costs of trades that fail due to a lack of communication are less than in the case of outright fraud or dishonesty. But when a buyer has won an auction, the seller has to pay a small commission on the sale, and may have other costs associated with listing the item. If there is no communication, or payment is not made, this is lost. In addition to the financial cost is an emotional cost in terms of the frustration of efforts to communicate with a trader that does not respond, and the loss of an opportunity to sell the item to other bidders.

As previously noted, the providers of online auction service provide guidelines that recommend using the available information, but these guidelines are minimal, so traders must develop their own strategies for how to use this information to make a decision as to whether or not to trust a seller. In the absence of specific advice on how to use the available information, buyers are left with a number of possibilities.

At the most extreme, a trader might decide to trust no seller, and bid on nothing, or to trust all traders and place bids on desired items, regardless of the history. Between these, traders can try to find optimal strategies through individual trial and error, participating in trades and trying to identify patterns in trades that go wrong. We would expect that a strategy developed in this way would be slow and costly to develop.

An alternative is to extend the sources of experience, by combining personal experience and drawing on the experience of others, in a form of social learning. Traders cannot discern much about another's strategy simply from following their trades. With a two step trading decision, if no bid is made, a potential buyer cannot tell whether this is because others are not interested, or because they don't trust the trader. A more likely path is that traders communicate elements of their strategy to each other, that is, strategy elements are directly communicated between traders. People are most likely to share this sort of information with those with whom they are friendly. In the model, traders are assumed to have friendly, information sharing, relationships with those with whom they have concluded successful trades.

The Trade Me system auction process

Typically, participants in online trading act both as buyers and as sellers at various times. The environment is clearly defined in terms of its scope, population, and the information displayed. The bounds of the environment are defined by the website, and the signed up membership. Members may be widely dispersed geographically, but are closely located in terms of the environment. All link to the website environment directly, and in exactly the same way.

Information on buyers and sellers is provided to traders only through the online service, and the information that is provided is very restricted. Trade Me provides information on geographical location, on the total number of trades, ratings of individual trades (as positive, neutral, or negative), links recording the goods sold, and comments that often provide more detailed information about any problems that were encountered. Most commonly, negative or neutral comments report failures to communicate, failures to go through with the purchase, and failures to deliver the goods as described, or a failure to deliver any goods at all.

A sale transaction begins with a seller offering an item for sale. Potential buyers decide whether to bid, and if they do bid, enter a bid price. The sale proceeds as a conventional auction, with the highest bidder at the close of the auction winning the auction. The people participating in trades remain anonymous until a sale agreement is concluded, until which they are identified only by their chosen nickname. At the conclusion of an auction, Trade Me provides buyer and seller with contact information. The buyer contacts the seller to arrange payment and delivery or collection of the goods.

The transaction can fail at any of these steps: the buyer can fail to make contact; the seller can fail to respond, the buyer can fail to pay, the seller can fail to deliver. While any of these outcomes are undesirable outcomes, only two have an actual cost. Where the buyer fails to pay, the seller is still liable for the selling fee. Where the seller fails to deliver, the buyer loses the dollar value of the agreed price.

The buyer is the most exposed to dishonest sellers, because it is common for goods to be paid for before delivery. For the buyer, in particular, a decision as to whether to bid involves an assessment of the trustworthiness of the seller. As previously stated, the information that is available before a bid is made is minimal. As a result, many of the cues that people might normally use to make assessments of other people, and decisions about whether or not to trust them are not present.

The services are relatively new, so the environment is a novel situation, particularly for people using these services for the first time. Although online trading sites recommend checking feedback information this advice is not especially prominent. eBay (nd) is specific about interpreting the information provided, while Trade Me (nd) suggests that buyers

> 'Review the member's selling feedback.... This shows other items they've sold recently and comments from previous buyers. Sellers may also want to view the member profiles of bidders to see how reliable they have been in past transactions. Usually, a high feedback score and high percentage is a good sign, but you should always check your trading partner's member profile to read comments and look for negative remarks.'

Even with the detail provided, naive traders do not necessarily know how to use this information to make a decision as to whether to bid. More experienced traders look for patterns in trading history before deciding whether to bid on an item. The development of these trading strategies of more experienced traders is not entirely a matter of trial and error, as traders share information on their own strategies and may modify their strategies in the light of the experiences of other traders. As sites like eBay and Trade Me have only been operating for a few years, buyers have developed strategies in a relatively short time. In that period, buyers have identified reputation and previous dealings as the most useful information in assessing a prospective trader's trustworthiness and reliability (Strader & Ramaswami, 2002).

Method: Constructing the Trading Model

In the Trading Model, the agents are provided with similar information to that provided by New Zealand's major online auction web site, Trade Me (www.trademe.co.nz). In the model, agents act as buyers and sellers in a simulated online trading market. The simulation provides potential buyers with a similar set of information to that provided by New Zealand's Trade Me internet auction web site (Trade Me, 2007). These elements of information provided are: the number of completed trades; the number of successful (good) trades; the number of times that the seller has failed to communicate; failed to pay; and failed to deliver. As trading progresses, a history is collated of the outcomes of each trade.

Each Trader begins with a randomly determined strategy, elements of which they can share with other agents. As the simulation runs, it models the dissemination of trust strategy information in a population of trader agents. Trust strategy information may be exchanged between traders that have concluded a successful trade, with the less successful trader randomly adopting elements of the more successful trader's strategy. This model can be used to explore what might happen when friendly agents can communicate information on trust strategies, and whether effective trust strategies might be adopted throughout the population through this process. If such a strategy is effective, traders should be able to discern trustworthy from untrustworthy agents. They should continue to trade with trustworthy agents, and decline to trade with untrustworthy agents. Untrustworthy agents should have less success selling, the proportion of bad trades should fall, and the overall happiness of should improve as the happiness cost of bad trades falls. If trust is operating effectively, trades with trustworthy agents should continue, and the happiness gains accrued from these trades should continue.

This model takes quite a different approach to modelling trust to the Basic Breaking Model described in the previous chapter above. Where that model was based on a very formal laboratory situation, this is a more recognisable real-world situation. In line with this, the psychological theory drawn on for the model is not situation specific. Instead of beginning with formal and very situation specific algorithm, the agents in the Trading Model simulation are modelled in terms of the information that is available to them, and possible ways that they might attempt to adapt to new information as they accumulate experience.

The model assumes that the decision-making process is a basic summation of the information that is available, with different pieces of information weighted differently. Agents begin with a randomly generated strategy, that they can modify as they adapt in response to experience.

In this situation, learning through purely individual trial and error would prove an expensive process. Further, observation would suggest that people share strategy tips with friends. For the model, adaptation is assumed to occur through a process of observational learning, whereby agents communicate some information on their successful strategies to each other. In the real world, that information tends to be communicated by friends. In the model, friendly relations are assumed following a successful trade, so agents may share information on their strategies with other traders that they have traded successfully with.

The model therefore models agents with a decision-making strategy for whether to trust or not. This strategy is refined through an observational learning process, as agents share information on successful strategies. The model has two components: the agents (Trader) and the main body of the programme (TradeMe)

Trader (agent)

Trader agents have randomly assigned reliability and honesty characteristics. These are fixed traits. These are used to calculate the chances of a trader failing to communicate (reliability), pay (reliability), or deliver goods (honesty). The trader agents have a number of actions that they can carry out: a new agent can initialise itself; it can decide whether it has something for sale; it can decide whether to bid, and how much to bid; it can decide whether to go through with the deal; and whether to adopt elements of strategy from its trading partner.

Agent initialises itself: init()

This initialises the randomly determined characteristics of each trader. This sets the fixed values for reliability and honesty, and the initial values for the weightings for each of bad communications, bad payments, bad deliveries, (all of which are failed trades), good trades and the total number of trades.

Agent decides whether it has something to sell: forSalePrice()

This action randomly generates a decision as to whether to offer something for sale. Agents have a 50% probability of offering something for sale. If the agent is selling something, this action also randomly generates an asking price.

Agent decides whether, and how much it wants to bid: bidToBuy (badDeliveries, badCommunications, badPayment, goodTrades, totalTrades)

The TradeMe environment passes five pieces of information to the buyer: badDeliveries; badCommunications; badPayment; goodTrades; totalTrades. The buyer calculates an assessment of the seller, using a simple summation of the weighted information.

$Assess = tT \times WeightTT + bC \times WeightBC + bP \times WeightBP + bD \times WeightBD + gT \times WeightGT$

Where

tT is the total number of trades that the trader has been engaged in,

bC is the proportion of bad communications

bP is the proportion of bad payments

bD is the proportion of bad deliveries

gT is the proportion of good trades

$WeightTT$ is the weight applied to total trades

$WeightBC$ is the weight applied to the proportion of bad communications

$WeightBP$ is the weight applied to the proportion of bad payments

$WeightBD$ is the weight applied to the proportion of bad deliveries

$WeightGT$ is the weight applied to the proportion of good trades

If the assessment is negative, the trader will not consider bidding. If the assessment is positive, the trader will randomly generate a decision whether to bid, and the bid price. The buyer also has a 50% probability of bidding.

Agent decides whether it is going to go through with the deal: communicate(), payUp(), delivery()

These actions return the agent's decision whether or not to communicate, to pay, or to deliver respectively.

Agent decides whether it will adopt elements of others' strategy: exchangeStrategy(Trader, partner)

This action occurs following the completion of a satisfactory trade. If the trader's overall happiness is less than that of their partner in a transaction, the trader collects the information on the partner's strategy. The trader then decides, element by element, whether or not to adopt that element from the partner's strategy. There is a 50% chance that the trader will adopt any one element.

TradeMe (the main body of the programme)

This section carries out two functions. The first function is that the TradeMe section forms the environment in which the trader agents operate. In this case, the model simulates an explicitly identifiable environment, the Trade Me system. The first function is that it manages the agent-based model house-keeping. This consists the initial set up of the model, scheduling the actions, running the simulation, and collecting the data and writing it to a file. The model has three actions scheduled, running a simulation iteration, or step, called TradeStep, outputting data, and stopping the simulation.

Part of the process of setting up the TradeMe model is initialising the population, and generating each of the agents and initialising the agents to generate their individual characteristics. An inherent part of this process is the decision about the number of agents in the population. The model needed enough agents to have a large range of different initial strategies, and enough agents to credibly simulate a market full of strangers. On the other hand, the population of the model needs to be small enough that the simulation runs in a reasonable time, and small enough that the data files generated are manageable for subsequent analysis. The population size of 100 was chosen as the data files and analysis were still manageable.

TradeStep() and sale()

The main component of TradeStep is the sale() function. This generates a round of sales. The action first shuffles the agents, to randomise the order in which agents are polled to see if they have anything for sale. One by one, the agents are asked to run forSalePrice(), then all of the other agents are asked to run bidToBuy(). The top bidder becomes the buyer. The programme then alternates polling the bidder for communications, the seller for communication, the bidder for whether they will pay, and the seller for whether they will deliver. If the transaction fails at any stage, the result is recorded against the defaulter, and happiness scores are allocated to record satisfaction with the result. Table shows the scores allocated for each of the types of outcome. Both buyer and seller are happy with a good sale outcome. This is reflected in a happiness score of 10 for a good trade. If the deal founders through bad communication by one party, the other will be somewhat unhappy, and gets a happiness score of -2. If the buyer does not pay, the seller will be a little more unhappy, and gets a happiness score of -5. In both situations the defaulter is presumably neither happy nor unhappy, and will get a neutral result. If the buyer pays, but the seller fails to deliver, the buyer is likely to be very unhappy. The seller is likely to be very happy with this outcome. The buyer gets a score of -20, and the seller gets a score of 20. Table 8.1 shows the tables with the outcomes for

Table 8.1: *Happiness scores for each possible outcome*

	Seller happiness	Buyer happiness
Buyer does not communicate	-2	0
Seller does not communicate	0	-2
Buyer does not pay	-5	0
Seller does not deliver	20	-20
Good trade	10	10

both seller and buyers.

Following the trading round the trader and partner may exchange information. This information exchange is in one direction only, from the trader with the greater total happiness score to the trader with the lesser total happiness score. This is intended to represent the less successful traders learning from the more successful traders. This exchange only occurs when the traders have negotiated a good sale. For each of the five weighting components, the trader with the lower happiness score has a 50% chance of adopting the weighting, so randomly adopts between zero and five of the weightings.

This information sharing is similar to the process used in genetic algorithms. Genetic algorithms are effective techniques for searching for optimum strategies. Their development was inspired by the molecular biological model of genetics. Genetic algorithms have been applied to psychological models, but in this setting their application has been less explicit in terms of what the genetic elements represent. This model adopts one feature of the genetic algorithm technique, the transfer of discrete chunks of information.

The trading model is intended to explore the role that observational learning might have in traders finding an optimal way of using the information that they have available on other traders. This form of learning has some features in common with genetic algorithms, notably the transfer of information among successful traders. There are, however, some important differences between the information exchange in the trading model and the way that information is combined in genetic algorithms. Firstly, while information is shared between traders who have had a successful interaction, information exchange is not restricted to being only between the most successful traders. In the early rounds, the most successful traders may be the most dishonest traders, as they can make substantial gains from each successful delivery default. But information is only exchanged following

Table 8.2: *Agent characteristics after initialisation round.*

			Weights given to scores				
	Characteristics		Failures to			No. of trades	
Agent	Reliable	Honest	Comm.	Deliver	Pay	Good	Total
1	0.78	0.86	0.81	0.22	0.41	0.65	0.0060
2	0.79	0.69	0.75	0.84	0.97	0.65	0.0062
3	0.84	0.67	0.35	0.65	0.03	0.86	0.0028
4	0.71	1.00	0.09	0.72	0.76	0.15	0.0068
5	0.53	0.80	0.39	0.95	0.17	0.52	0.0027
6	0.37	0.97	0.47	0.89	0.54	0.54	0.0027

a successful trade. Secondly, there is only one generation in this model, whereas in genetic algorithms, information is transferred from one generation of agents to the next.

Details of a single trading round

Computer simulation allows us to collect every element of detail at every step. While this would generate an overwhelming amount of data if done for an entire set of simulation runs, it does allow us to see each decision and interaction for a small number of traders, over a small number of trading rounds. Carrying out an inspection of these has a number of benefits. A step by step inspection of the program in action can help provide some assurance that it is doing what it should be doing, and so forms part of the verification process. It can also provide some hints as to how to proceed in the modelling. For example, in some runs strategies that were very sensitive to the total number of trades developed early in the simulation. As a result, with few previous trading rounds, none of the traders had enough trades for any of the others to trust them and trading stopped.

The results of the initialisation round is shown in Table 8.2. At initialisation, two of the agents (Agent 5 and Agent 6) have very low reliability, and two other agents (Agent 2 and Agent 3) have moderately low honesty.

Following the initialisation round, the model carries out a series of trading rounds. Each agent is polled as to whether it has anything to sell. The order in which the agents are polled is generated randomly, using one of the functions of the Repast package. For the agent being polled, the decision is effectively a coin toss; if a random number generated from a uniform distribution with a range zero to one (Uniform[0,1]) is greater than 0.5, then the seller has something to sell, otherwise it doesn't. Effectively, this is the equivalent of a 50/50 coin toss. The auction round begins with Agent 2 deciding whether it has something to sell.

Agent 2 generates a random number (0.57) that is greater than 0.5 so agent 2 has something for sale.

If an agent does have something to sell, the bidding round follows immediately after the decision to sell. In the bidding round, each agent is polled as to whether it wants to bid. There are three steps in each agent's decision as to whether and how much to bid. First it decides, based on the seller's history, and its own strategy, whether it trusts the agent. Next, if the bidder does trust the seller, it decides whether or not it is going to bid. Again this decision is effectively a coin toss. If the agent does decide to bid, the amount bid is randomly generated from a normal distribution with a mean of 20, and a standard deviation of 5.

At this stage, there have been no trades, and so none of the agents have a trading history. Agents 1, 3, 4, 5, and 6 all decide that they will trust the seller, and move on to the second step in the decision-making process. Only one of the agents, Agent 5, decides to bid, and it bids 30.35.

The auction concludes with the highest (and only) bidder, Agent 5 winning the auction. With a sale agreement reached, the two agents enter a phase in which they attempt to complete the transaction.

The first step requires the buyer to initiate contact with the seller. On Trade Me, this step provides the seller with the buyer's contact details and delivery address. Whether this happens is determined randomly, with the decision depending on the random number and the agent's reliability. The agent generates a random number from a Uniform[0,1] distribution. If the number is greater than the agent's reliability, the agent fails to communicate. For example, a very reliable agent, with a reliability of 0.95, would have to generate a random number between 0.95 and 1.0 for a failure of communication to happen. Agent 5 is less reliable, and if it generates a random number between 0.534 and 1.0, it will fail to communicate.

Agent 5 generates a random number (0.50) that is less than its reliability (0.534), and does communicate with the seller.

The next step requires the seller to return the communication. On Trade Me, this would typically involve the seller providing an account number to enable the buyer paying. This is determined in the same way that the buyer's communication is determined.

Table 8.3: *Agent 2's use of Agent 5's history in making a trust decision.*

	Weights given to scores				
	Failures to			No. of trades	
Agent	Comm.	Deliver	Pay	Good	Total
2	0.75	0.84	0.97	0.65	0.0062
	Seller's history				
	Failures to			No. of trades	
Agent	Comm.	Deliver	Pay	Good	Total
5	1	0	0	0	1
	Decision result				
Assess	Comm.	Deliver	Pay	Good	Total
-0.744	-0.75	0	NA	0	0.0062
Do not trust					

Agent 2 generates a random number (0.35) that is less than its reliability (0.79), and does return the communication.

With communication established, the next step in the sale process requires the buyer to pay. There is no gain for the buyer in failing to make payment, as the seller does not deliver until the buyer has made payment. Again, this decision as to whether to pay depends on the reliability of the buyer, and is made in the same way.

Agent 5 generates a random number (0.65) that is greater than its reliability (0.534), and does not pay up. The sale process stops at this point. Agent 2 is unhappy about the outcome, and its happiness drops 5 points. Agent 5 is indifferent to the outcome, so its happiness remains unchanged. A bad communication is recorded on Agent 5's trading record, and both Agent 2 and Agent 5 have a trade added to their total trades record.

In the next trading round, it happens that Trader 5 has something to sell. At this point, Trader 5 has no good trades, and one trade that has failed because they did not communicate - a 100% failure rate. At this point, All of the traders assess Trader 5 as being untrustworthy. The details of the trust assessment calculation for Trader 2 is shown in Table 8.3.

Where a trade is completed with no failures, the agents compare happiness scores. The agent with the lowest score carries out the equivalent of a coin toss for each element of the trust decision strategy. An example of the process is shown in Table 8.4

Table 8.4: *Adoption of strategy elements by Agent 2.*

		Weights given to scores				
		Failures to			No. of trades	
Agent	Happiness	Comm.	Deliver	Pay	Good	Total
1	0	0.81	0.22	0.41	0.65	0.0060
2	-7	0.75	0.84	0.97	0.65	0.0062
Agent 2 randomly adopts elements of Agent 1's strategy						
coin toss		1	0	1	0	0
New strategy		0.81	0.84	0.41	0.65	0.0062

Results

Data were collected from 25 different runs of the Trading Model, each of which consisted 250 auction trading rounds. Experience working with the model suggested that in most cases the simulation had settled to a final distribution of strategies by about round 250. It is likely that this is driven by the dynamics of the model, as these determine the rate at which traders exchange strategy information with each other. A new population of traders was generated for each run, for each of which honesty and reliability characteristics were randomly generated, as were an initial set of strategy weightings. Other model parameters, such as the values placed on each type of outcome, were fixed across all 25 runs.

The results were analysed using the R (Ihaka & Gentleman, 1996) statistics package.

Dimensions

This model tracks the activities of a population of agents operating in a new market. Each iteration of the model corresponds to a single round of trading in this market. This generates longitudinal population data, that can be sliced at any step to produce cross-sectional data. A single run of a number of trading rounds also generates a trading history for each agent. Finally, the model can be run a number of times, each of which run might correspond to a new population starting trading in a new market. Figure 8.1 shows the structure of the data collected from the model. The different populations, starting conditions, and trading histories lead to different possibilities as to how a market might develop. And with data collected for every individual, at every trading round, across a number different runs, the model generates data in a number of different dimensions.

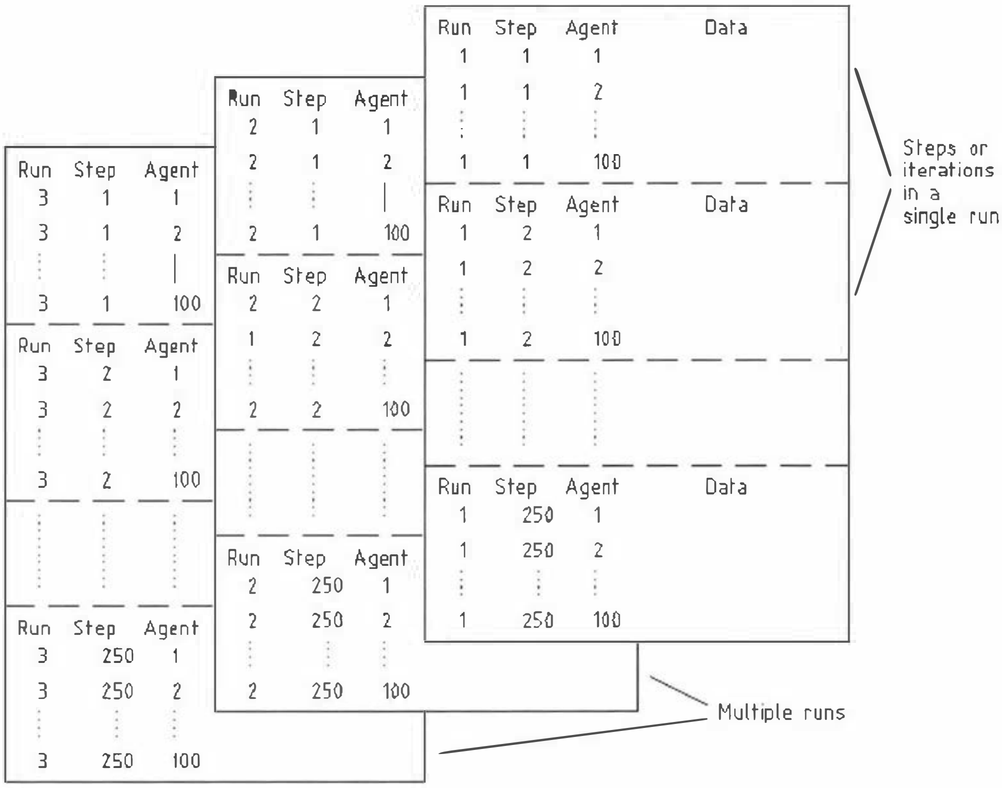

Figure 8.1: Structure of results data output by the Trading Model

At a population level, the model might generate patterns that appear across all runs. An example is that the design of this model is such that we would expect that the agents reach a consensus on the weightings given the strategy elements. This would be expected because there is no innovation in trading strategy, and the direction of transmission is asymmetric; strategy elements are always passed from more successful to less successful traders.

An agent-based model can, in general, be used to investigate two related aspects of the behaviour of a network of traders: the stable patterns arising within the online market, and the dynamic behaviour of traders in the system. In the case of the Trading model, the agent-based model can be used to explore the way that strategic information spreads through the population of traders under a particular set of assumptions. It is less clear whether passing information is allowing a population of traders to maintain trading with trustworthy traders, and to identify others.

The model can also be used to observe the development of strategy patterns in time. Sequential data can be collected at a number of intervals throughout a run. In the case of the Trading model, each run of the model consisted 250 trading rounds, with data collected at each trading round. The collection of data at every trading round provides a longitudinal history about individual histories, and about how the system has behaved over time. Thus we can collect information on the dynamic behaviour of the system.

The extent to which resulting strategies are sensitive to individual trader characteristics, and to unfolding trading interactions can be explored through multiple runs of the model. It is possible that common patterns develop across all runs. Alternatively, it is possible that there are a spread of possible outcomes, some of which may be more likely than others.

Model validation

The first step, before analysing the data, was to check that the model was behaving as I expected it to, that is, in accordance with its design. A single run was used for this verification.

In the model, reliability and honesty are generated by drawing random numbers from a simulated uniform distribution, then taking the fourth root of these. The intention was that this would generate a distribution of reliability and honesty with a large negative skew: a population that was largely reliable and honest.

The untransformed distributions of honesty and reliability are shown in Figure 8.2 . Both were negatively skewed. This indicates that, as intended, the population has a small number of agents with low honesty, and a small number of agents with low reliability. The majority of the agent population is reliable and honest.

The chance of bad communications or bad payments occurring during a trade is driven by the reliability of the trader. This should lead to a very strong correlation between each of these behaviours and reliability. The number of trades that fail through bad communications or bad payments should also feed through into the number of good trades, as each failed trade reduces the possible number of good trades. I expected to find a positive correlation between the number of good trades and trader reliability. While the model is such that there is no happiness cost for each bad communication or bad payment, each does result in an opportunity cost, the loss of the potential happiness resulting from a completed good trade. As a result, I expected that trader reliability would also be inversely correlated with happiness.

The structure plot matrix (SPLOM) at the 250th round (Figure 8.2) shows that, as expected, there were strong negative correlations between the

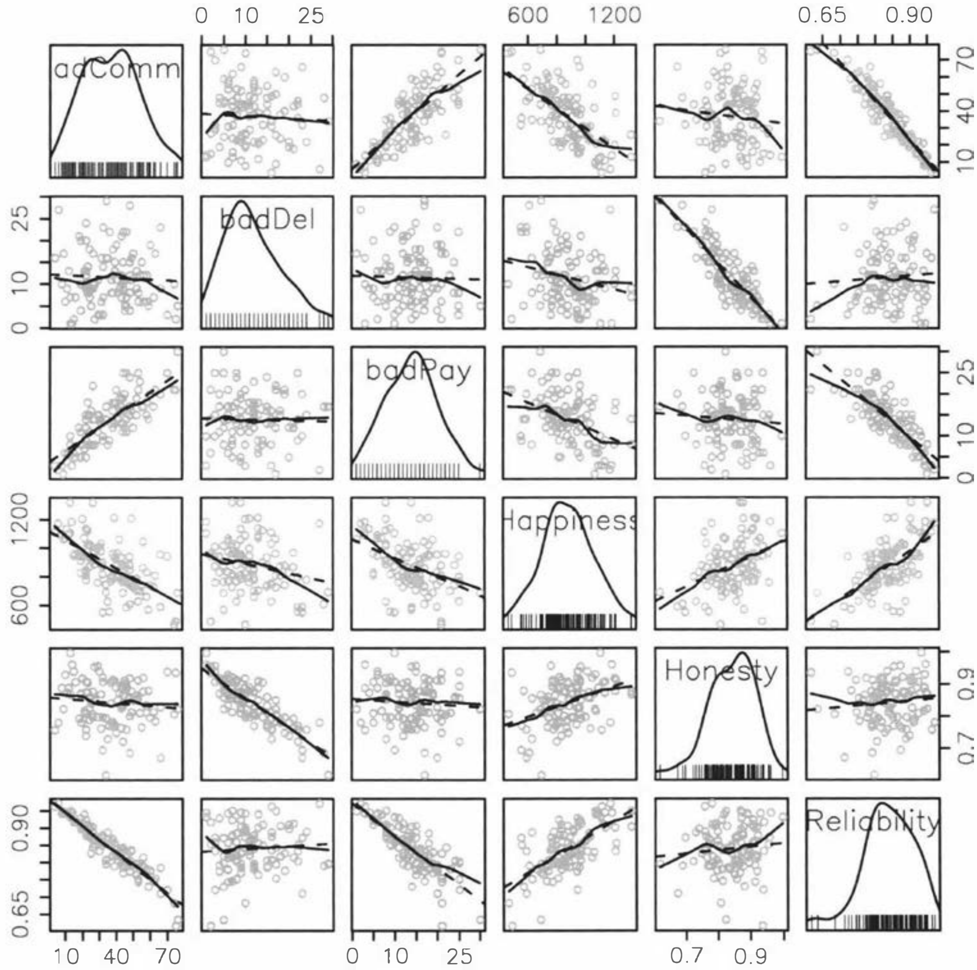

Figure 8.2: Structure plot matrices showing correlations between the outcome variables

number of bad communications and reliability, $r(100)=-.939$, $p<.001$, and a slightly weaker correlation between the number of bad payments and reliability, $r(100)=-.804$, $p<.001$. There was also a strong positive correlation between reliability and the number of good trades, $r(100)=.820$, $p<.001$, and a weaker positive correlation between reliability and happiness, $r(100)=.645$, $p<.001$. As expected, the number of trades that failed through bad communications and bad payments was strongly negatively correlated with the traders' reliability, and positively correlated with the number of good trades completed by the trader, and this difference in the number of completed good trades was reflected in the positive correlation between happiness and reliability.

The model was also written so that the chance of a bad delivery depends on the trader's honesty. I expected to find a strong negative correlation between honesty and the number of bad deliveries. The SPLOM (Figure 8.2) shows a strong negative correlation between the number of bad payments and the reliability of the trader, $r(100)=-.842$, $p<.001$. Again the number of good trades was reduced by the number of trades that fail for a bad payment, so good trades were correlated positively with honesty, $r(100)= .354$, $p<.001$, and, in turn, with happiness $r(100)= .407$, $p<.001$.

Exploration of data generated by the model

Potential for capture of strategy by dishonest traders

The agents judge the relative success of their strategies by comparing their happiness with that of other traders with whom they have completed a successful trade. The trader who is less happy may adopt strategy elements from the more successful trader.

The initial model design was based on the assumption that happiness gains made from both successful trades and from successful frauds would both count when the agents compared strategies. As a successful fraud carries a high reward, it was possible that the strategy might become captured by less honest traders. If these traders can make large gains from fraud than from legitimate trading, they may become the happiest traders, and become influential in disseminating strategy because of their happiness. For example, a strategy that benefits dishonest traders might include trusting traders despite a history of failed deliveries.

The model was run 25 times under the condition that the gains from fraud

Table 8.5: *Per round mean and median values at the 250th trading round.*

	Gains from fraud not included		Gains from fraud included	
	M (*SD*)	Median	*M* (*SD*)	Median
Median happiness	3.256(.201)	3.280	4.238(0.206)	4.216
Total happiness	325.7(19.3)	328.6	425.8(18.5)	427.0
Bad communications	15.62(1.66)	15.61	15.37(0.82)	15.30
Bad payments	5.612(0.235)	5.628	5.517(0.228)	5.556
Bad deliveries	4.788(0.332)	4.808	4.874(0.264)	4.860
Good trades	48.07(1.65)	48.36	48.42(1.62)	48.43

are included in the happiness score, and 25 times under the condition that the gains from fraud are not included in the happiness score. Each run represents one single possible history and set of outcomes in a population.

The final (round 250) per round mean and median values for happiness scores and for the types of outcome (bad communications, bad payments, bad deliveries, and good trades) is shown in Table 8.5. The overall happiness was higher under the condition that the gains from fraud (bad deliveries) were included, reflecting the inclusion of these gains in the overall happiness figures. The numbers of each type of outcome remained the same whether or not gains from fraud were included. Most notably, including the gains from frauds did not result in a difference in the overall number of bad delivery outcomes. This suggested that the least honest traders did not influence strategies so that other agents would trust dishonest traders. Subsequent analyses were carried out only for the model in which the gains from frauds were included in the happiness outcomes.

Analysis across all 25 runs of the model

Early and late outcomes Possibly the most basic question that could be asked of the data is whether there is a shift in the outcomes throughout a simulation run. This question might be asked of the data from a single run, but there is no assurance that this is not just one possible outcome, rather than being a sustained pattern, or a likely outcome. The change in outcomes across all 25 runs is shown in Table 8.6 . The only outcome showing a significant change across the whole population is the median happiness, with all other outcomes showing small non-significant changes.

Table 8.6: *Difference in overall per round outcomes between round 5 and round 250 across 25 runs*

Outcome	Difference	$t(24)$	p
median happiness	0.3936	3.076	.0052
total happiness	14.94	1.538	.137
bad communications	0.0818	0.235	.817
bad payments	-0.1143	-0.554	.585
bad deliveries	-0.2742	-1.563	.259
good trades	-0.8892	0.997	.329

Consensus on trust strategy In 25 runs of the model, the population of agents found a consensus on all weightings only in one run. Other runs reached varying degrees of consensus, from no consensus to consensus on all five weightings (see Figure 8.3). Even where consensus on strategies is not reached, the number of different strategies being used in the population reduces. Overall, across 25 runs, and five strategies, in the great majority of cases, the strategies being used in the population had reduced either to a consensus or to two different weightings (see Figure 8.4). The model behaves as expected, with strategy weightings gradually moving toward consensus over a number of runs.

Clustering of runs

The 25 runs produced a variety of different outcomes and weightings by the 250th trading round. The number of dimensions made it difficult to detect whether there were patterns in these outcomes. Clustering techniques are useful for identifying runs that are similar. It does this by identifying which groups of elements are closest to each other, and identifying these with clusters. If the clusters are differentiated, there will be a distance between the clusters. K-means clustering begins with a set of randomly determined centroids , with each run allocated to the cluster with the nearest centroid. Once all runs have been allocated to a cluster, the centroid is recalculated. The process continues with reallocation, and recalculating of the centroids until there are no more changes in allocation. A clustering analysis was carried out on the 25 runs, in an attempt to identify whether sub-groupings of outcomes had formed, and whether these were related either to the weightings that were being applied by the agents by the end of a simulation run, or to the values of the outcome variables early in the simulation run.

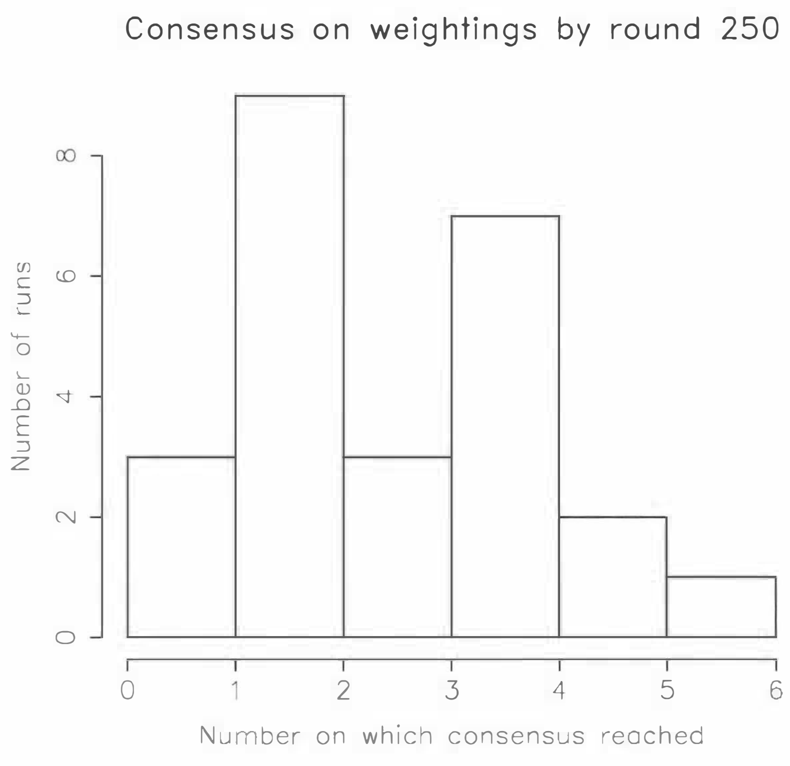

Figure 8.3: Distribution of the number of times consensus is reached after 250 rounds, across 25 simulation runs. Consensus can be reached for a number of strategies, between between no consensus on strategies and consensus on all five strategy elements.

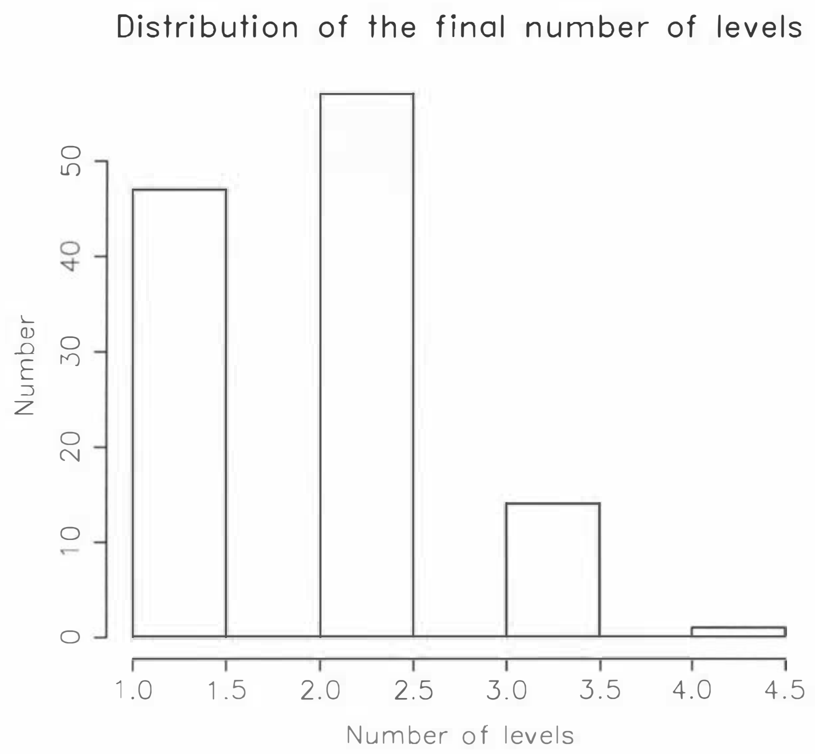

Figure 8.4: Distribution of the final number of strategy levels at round 250, across 25 runs, each with five different strategies. Most of the strategies have reduced so that either the population has reached a consensus on the strategies, or to there being two different strategies in the population.

Given that cluster analysis is an attempt to find which runs are similar, a prerequisite step requires decisions about what things might be similar, the clustering variables, and about what defines similarity. In k-means clustering, similarity is determined by distance to the cluster centroid. For this analysis there were a number of possible clustering variables. The primary marker of outcome was a single variable, happiness. Happiness is directly derived from trading outcomes, with trades that fail, and trades that succeed, all adding happiness or unhappiness to the overall total. Using happiness directly as the clustering variable would not differentiate these underlying sources of this happiness outcome.

One alternative was to use the trading outcomes that contribute to generating the happiness, and unhappiness, that aggregates into a final happiness outcome. There are four possible outcomes of a trade: a failure to communicate, a failure to pay, a failure to deliver, and a completed good trade. Of these, the number of good trades was strongly correlated with the number of failures to communicate, $r(23)=-.860$, $p < 0.00$). This suggests that the *good trades* outcome and the *bad communications* strategy outcomes were similar enough that including both in the clustering would amount to double-counting the number of bad communications. This left three outcome variables in the clustering: bad communications, bad payments, and bad deliveries.

Clustering was carried out using k-means clustering on the per round occurrence of three outcome variables in the whole population: bad communications; bad payments; and bad deliveries. As the accumulated number of bad trades at the 250th round is inherently larger than the accumulated number of bad trades at the 5th round, the raw number of failures was divided by the respective number of rounds to obtain the rate of occurrence of the different outcomes, and thereby a comparable result for the early and late outcomes.

The plot of the sum of squares from the k-means clustering (see Figure 8.5) shows no obvious knee-points to guide the choice of the number of clusters. In the absence of a clear indication from the data, the number of clusters was selected to be three. This gave a combination of simplicity, reasonable cluster sizes, and a coherent story. The centroids of the three clusters are as shown in Table 8.7. Figure 8.6 shows the clusters located in the plane formed by the two major principal components. The three centroids were strongly differentiated in their location along the axis defined by the first component of the principal components analysis, while all three centroids

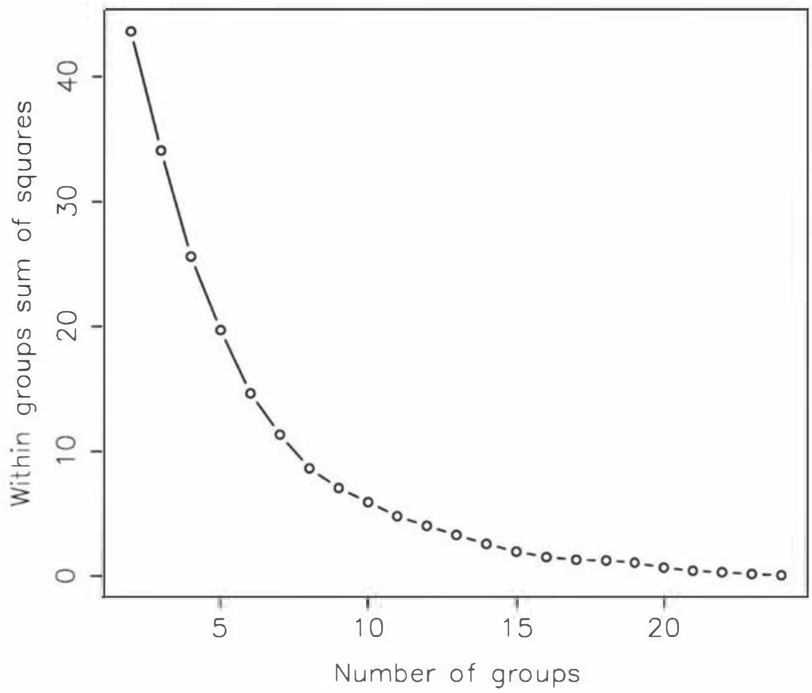

Figure 8.5: Weighted sum of squares for clusters

Table 8.7: *Cluster centroids for the three cluster result*

	Bad communications	Bad Deliveries	Bad Payments	Size
Cluster 1	1.47	-0.99	1.84	3
Cluster 2	-0.86	0.57	-0.74	10
Cluster 3	0.34	-0.23	0.16	12

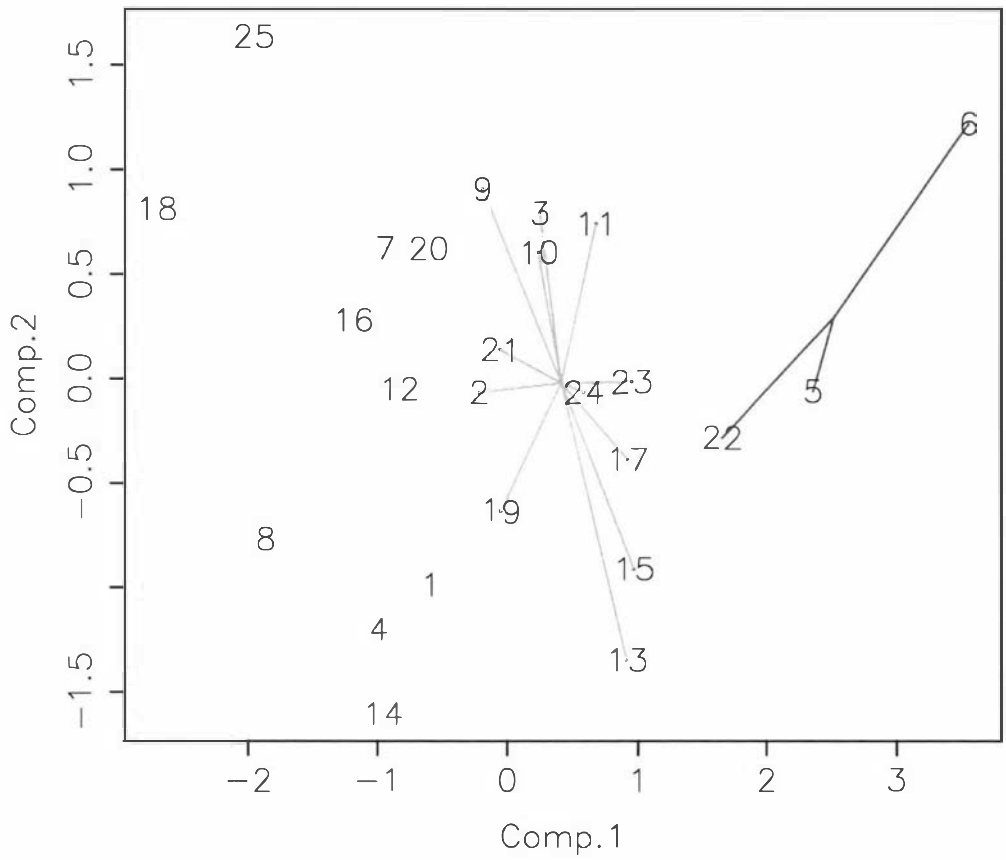

Figure 8.6: The three outcome clusters in principal-component space

were located near the mid-point of the second component. Bad communications and bad payments were loaded heavily on the first component, therefore this component reflects reliability driven trading failures. Bad deliveries, reflecting honesty driven trading failures, were loaded heavily on the second component. The distribution of the cluster centroids along the first component axis suggests that the clusters were differentiated in their numbers of reliability driven trading failures.

Of the three clusters, Cluster 1 was a small set of three runs in which the population suffered very high levels of trade failures due to unreliability (bad communications and bad payments), and very low levels of trade failures due to dishonesty (bad deliveries). Cluster 2 had the lowest levels of trade failures due to unreliability, and the highest level of trade failures due to dishonesty. Cluster 3 contains around half of the runs. It had a moderate level of failures due to unreliability, and a moderate level of failures due to dishonesty.

Outcomes by cluster Boxplots of the number of bad communications and bad payments at rounds 5 and 250 are shown in Figure 8.7 , alongside the corresponding boxplots of these outcomes at round 5. The difference between the early and late boxplots reflects the effect of strategies being transmitted and adopted through the population. For Clusters 2 and 3, the early and late outcomes are similar, but Cluster 1 showed a large increase in the number of failures due to bad communications between rounds 5 and 250.

Boxplots of the number of bad deliveries are shown in Figure 8.8. Again, Cluster 1 showed the biggest shift, with a large drop in the occurrence of bad deliveries.

Figure 8.8 also shows the numbers of good trades at rounds 5 and 250 of the simulation. Unlike the other types of outcome, the number of good trades was not directly used in differentiating the three clusters. The number of good trades is, however, directly restricted by these, as each failed trade represents a failed opportunity to complete a good trade. There is a clear difference between the three clusters in the number of good trades being completed at the end of the simulation. Cluster 1, with a high rate of reliability-driven trading failures and a low rate of dishonesty-driven trading failures, is the worst performing cluster in terms of good trades completed. Of the other two clusters, Cluster 2, has the highest number of good trades completed. This cluster has the higher rate of dishonesty-driven

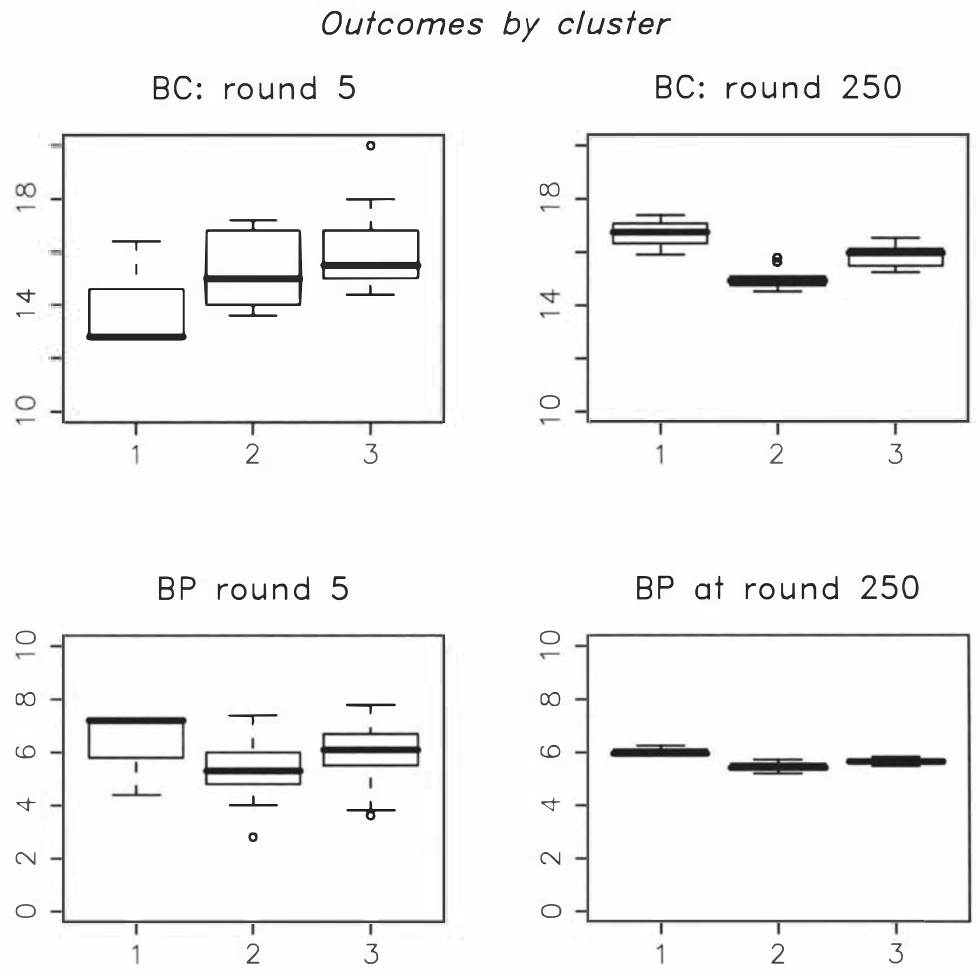

Figure 8.7: The number of bad communications (BC) and bad payments (BP) generated in each cluster early (trading round 5) and late (trading round 250) in the simulation runs

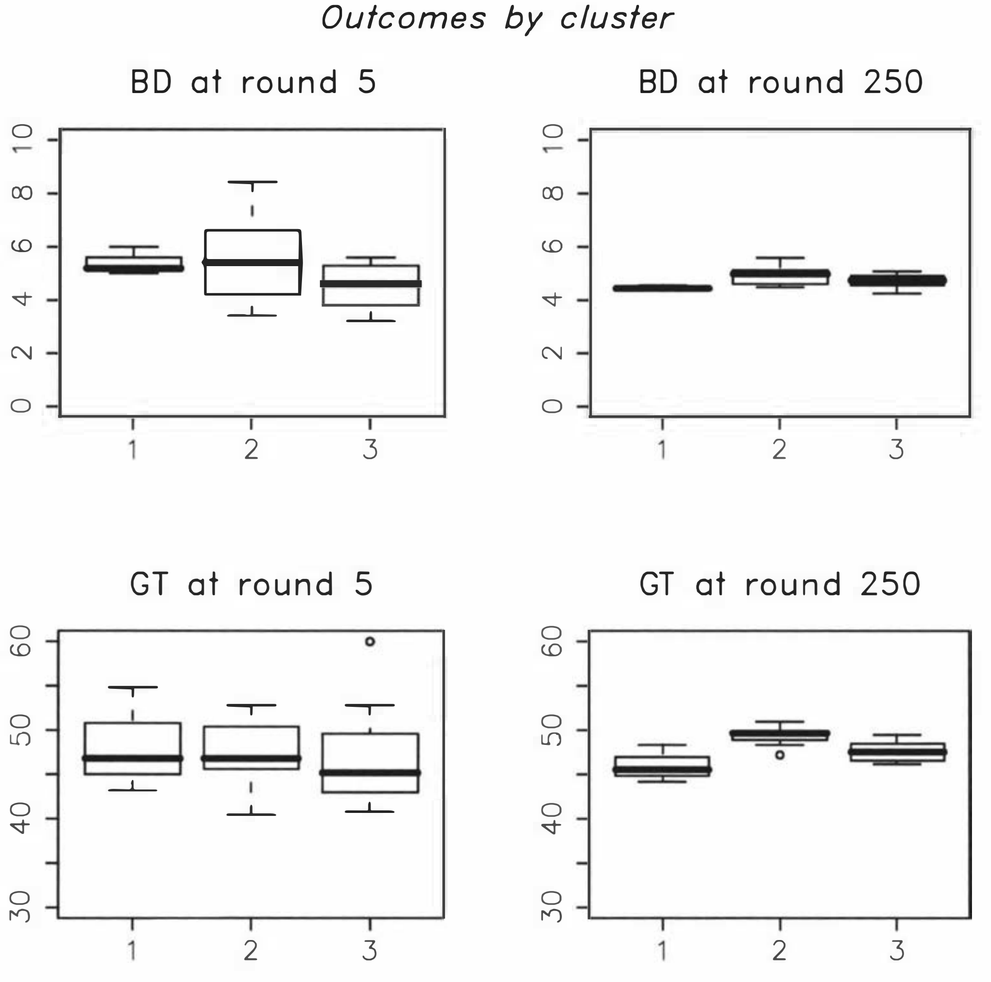

Figure 8.8: The number of bad deliveries (BD) and good trades (GT) generated in each cluster early (trading round 5) and late (trading round 250) in the simulation runs

trading failures, and a lower rate of reliability driven trading failures. The difference between the clusters in the number of good trades completed is significant, $F(2,22) = 10.75$, $p<.001$. There is a corresponding significant difference in the total happiness in the cluster populations, $F(2,22) = 4.18$, $p=.029$.

Across the three clusters, the form of the rate of occurrence of good trades is the mirror image of the patterns for bad communications and bad payments seen in Figure 8.7. This tends to suggest that bad communications and bad payments have the greatest influence on the number of good trades completed.

More surprisingly, the numbers of good trades parallels the numbers of bad deliveries. This is quite different from the mirror image pattern seen with bad communications and bad payments. This tends to suggest that bad deliveries do not translate so directly to a reduction in the number of good trades. It may be that while a decision not to trade with another trader may protect a buyer from a dishonest trader, it may also restrict the opportunities for good trades to be completed. A trade that fails due to bad delivery may be more than compensated for by other trades that are allowed to proceed, in the face of some evidence of previous dishonesty by the trader.

Weightings by cluster The clusterings were based on the outcomes, with no direct relationship between these and the weightings of the various strategy elements. A series of boxplots were also generated to investigate whether the outcomes might reflect patterns in the weightings on the strategy elements. Figure 8.9 shows the median weightings on bad communications and on bad payments, again at rounds 5 and 250. The weightings on bad communications are similar across all three clusters, but the three clusters have quite different median weightings on bad payments, with Cluster 1 showing a much lower weighting on this element, and Cluster 3 showing the highest weighting. There are larger differences apparent in the weightings on bad deliveries and on good trades. These are shown in the boxplots in Figure 8.10 . All three clusters have shown a reduction in the weighting on bad deliveries, with Clusters 1 and 3 showing the greatest drop, and lowest median values on this weighting. These two clusters have the highest weighting on the number of good trades. Cluster 2, which has generated the highest number of good trades, has the highest median weighting on bad deliveries, and lowest median weighting on good

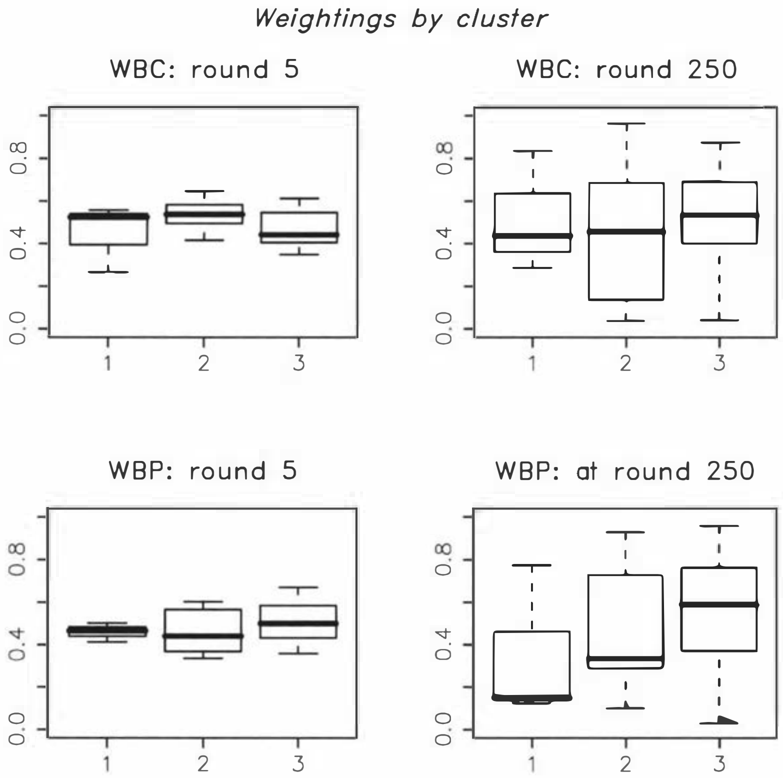

Figure 8.9: The strategy weightings on bad communications (WBC) and bad payments (WBP) in each cluster early (trading round 5) and late (trading round 250) in the simulation runs

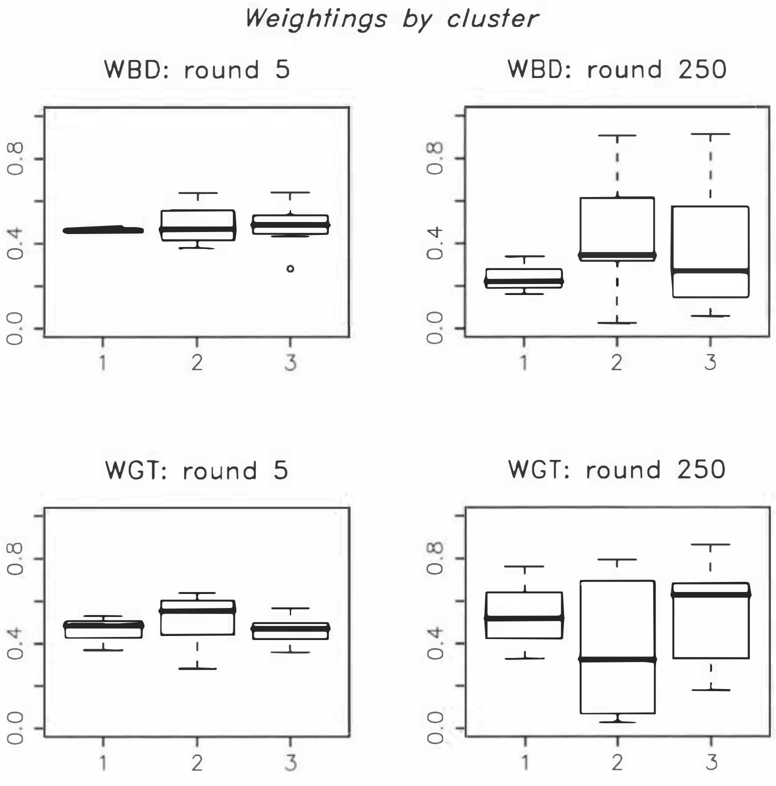

Figure 8.10: The strategy weightings on bad deliveries (WBD) and good trades (WGT) in each cluster early (trading round 5) and late (trading round 250) in the simulation runs

trades.

The boxplots of the weightings suggest that the difference in outcomes between the clusters may reflect a difference in the patterns of weightings in the clusters. Further, the median weightings have shifted from early to late in the cluster. This suggests that the three clusters may be following different patterns of evolution of the weightings.

Summary: Outcomes from many runs The previous analyses were carried out over a number of different runs, each of which might represent a number of possible ways that a market might evolve, or the evolution of different markets. This might be directly related to the real world equivalent of the development of different strategies within different online trading services with different populations. For example, the development of the strategies applied by the traders operating in eBay and Trade Me might be quite different, with that difference a result of nothing other than the whole population being drawn to different clustering outcomes as a result of the chance elements of initial conditions, individual characteristics, and the chance encounters of individual trades.

The results suggested that while there was little indication that the model generated any patterns at the highest level of aggregation, at the level of a number of runs, this did not mean that the outcomes were homogeneous. Reducing the level of aggregation a little, it is possible to identify possible clusters of run outcomes that share characteristics not in both the outcomes on which the clusters were based, but also in the consensus on strategies developed over time.

Working down in aggregation level, the next step might be to ask whether these patterns in clusters might be under-laid with patterns in the histories of development of the trust strategies in individual runs. In the next section, I look more closely at some individual runs of the model. The runs chosen for this level of analysis were the three runs closest to the centroids of the three clusters. These runs were: Run 5, closest to the centroid of Cluster 1; Run 12, closest to the centroid of Cluster 2; and Run 24, closest to the centroid of Cluster 3.

Zooming in: More detail from three runs representative of the three clusters

The cluster analysis carried over the full 25 runs generated a possible grouping of the runs into three clusters. These were defined primarily by the number of trading failures due to bad communications and to bad payments, that is trading failures resulting from agent unreliability. There are two distinct possibilities as to why trading failures might have differentiated in these clusters.

Firstly, there might have been chance differences in the reliability and honesty of agents in the agent populations of these different runs. The design of the model is such that any such differences would result in generating differences in the number of trading failures. This can be controlled for by statistically controlling for reliability and honesty.

Secondly, the clusters might have differed in the strategies that the agents have evolved. Any such strategy differences may have affected the number of trades being made as agents have decided whether or to to trade with a seller.

Change in the rate of occurrence of various outcome types If the development of strategies has the effect of allowing the population of traders to identify untrustworthy traders, I would expect a change in the rate of occurence of failed trades. The Poisson distribution describes the random occurrence of infrequent discrete events that occur at a fixed average rate. Poisson regression allows us to estimate the average rate of occurrence of the different types of outcome. In order to identify whether there was a difference between the clusters in the rate of the different types of outcome, I constructed a series of Poisson regression models. These modelled the number of each type of outcome (bad communications, bad payment, bad delivery, good trade), regressed on the cluster, and on reliability (bad communications, bad payment), or honesty (bad delivery), or both (good trade). Each cluster was represented by its representative run's value along the dimension defined by the first principal component. This was found during the cluster analysis, and was used in generating Figure 8.6, which shows the three clusters separated along this dimension. The weights on the first principle component were: 2.366 for Run 5; -0.8276 for Run 12; and 0.5921 for Run 24.

The results of these Poisson regressions are shown in Table 8.8 . This ta-

Table 8.8: *Poisson regression results for each type of outcome, regressed on the cluster principal component 1 value, reliability, and honesty across Runs 5, 12, and 24. The regression is carried out across all 250 trading rounds in each simulation run.*

	Estimate	p
Bad communications		
Intercept	1.43	<.001***
Value principle component 1 (cluster)	-0.021	.003**
Reliability	-4.02	<.001***
Bad payments		
Intercept	-0.511	<.001***
Value principle component 1 (cluster)	0.005	=0.689, n.s.
Reliability	-2.919	<.001***
Bad deliveries		
Intercept	0.613	<.001***
Value principle component 1 (cluster)	-0.021	=.109, n.s.
Honesty	-4.488	<.001***
Good trades		
Intercept	-2.92	<.001***
Value principle component 1 (cluster)	-0.014	<.001***
Reliability	0.709	<.001***
Honesty	1.915	<.001***

** significance < .005, *** significance <.001

Table 8.9: *Change in Poisson regression coefficients for each outcome type, regressed on Cluster, Reliability, and Honesty, between trading rounds 1-50 and 200-250.*

	Poisson regression coefficients		
Bad communications	Early rounds (1-50)	Late rounds (200-250)	Difference (95% *CI*)
Cluster	-0.021	-0.021	0.0008 (-0.0305, 0.0321)
Reliability	-4.180	-4.019	-0.162 (-0.483, 0.160)
Honesty	NA	NA	NA
Bad payments	Early rounds (1-50)	Late rounds (200-250)	Difference
Cluster	-0.010	-0.008	-0.0026 (-0.0556, 0.0503)
Reliability	-3.192	-3.061	-0.131 (-0.738, 0.477)
Honesty	NA	NA	NA
Bad deliveries	Early rounds (1-50)	Late rounds (200-250)	Difference
Cluster	0.030	-0.039	0.0692 (0.0137, 0.1246)*
Reliability	NA	NA	NA
Honesty	-4.550	-4.561	0.0117 (-0.703, 0.726)
Good trades	Early rounds (1-50)	Late rounds (200-250)	Difference Difference
Cluster	-0.024	-0.031	0.00718 (-0.0107, 0.0251)
Reliability	2.057	1.226	0.831 (0.533, 1.129)*
Honesty	0.726	-4.572	5.297 (4.990, 5.605)*

ble shows that, as expected, reliability was a significant predictor of bad communications and bad payments, honesty was a significant predictor of bad deliveries, and both reliability and honesty were significant predictors of the number of good trades. The cluster was a significant predictor of bad communications, and of good trades, but was not a significant predictor of bad payments or of bad deliveries. Cluster was not as strong a predictor as reliability and honesty, but these two characteristics were specifically included to drive the numbers of bad trades occurring. In contrast, a difference between clusters was not specifically designed into the model.

While this analysis showed that there was a difference between clusters in the rate of occurrence of bad communications and good trades, it does not show whether this was a result of a change in the rate of occurrence of each outcome, possibly as a result of the adoption of more or less effective strategies. This was investigated through a second set of Poisson regressions, carried out over the first and last 50 runs. The differences between these were calculated, and a confidence interval of the difference constructed. The results are shown in Table 8.9 . The results show that there was no change in the coefficients driving the rate of occurrence of bad communications, or bad payments. There was a significant change in the cluster coefficient for bad deliveries, such that the number of bad deliveries increases most in Cluster 1, less in Cluster 2, and least in Cluster 3.

Mean happiness in the three representative runs

The previous results have suggested that there are differences in outcomes from the simulation, and that these might be generating three clusters with similar outcome patterns. A key question in identifying differences in outcome is whether these differences reflect differences in the mean happiness in the population. There was a significant difference in the final mean happiness between Run 5, M = 761.58 , SD = 198.8768, Run 12, M = 834.31 , SD = 182.9162, and Run24, M = 808.64, SD = 194.9946, $F(297)=3.676$, $p < 0.05$. At least one run has generated an overall happiness that differs from the others significantly. Run 5 has generated an overall lower happiness than the other two runs.

Happiness outcome for individuals within the three representative runs

The analysis above suggests that there is a significant difference in outcome in the three representative runs, but provides no information on the distribution of happiness among the traders. Table 8.10 shows the changes in happiness of four trader agents over the first and last 25 runs, and the results of t-tests on the differences. The agents investigated were the most and least reliable traders, and the most and least honest traders. Run 12 has significant changes in the happiness of both the least reliable and the least honest traders.

Analysis of a single run (Run 12)

The data from the simulation run can provide some indication of the range of possible outcomes that might be generated in an online trading environment in which agents have little guidance as to how to apply information to making decisions about whether or not to trust another trader. It is also possible to look at the data collected from a single run. For example, it would be possible to more closely investigate what might happen in

Table 8.10: *Happiness of the most and least reliable traders, and the most and least honest traders for Runs 5, 12, and 24, early (Rounds 1-25) and late (Rounds 226-250) in the simulation*

	Mean Happiness			
	Early	Late	$t(48)$	p
Run 5				
most reliable	0.67	0.00	0.257	.78, ns
least reliable	0.92	1.84	-0.711	.48, ns
most honest	1.92	0.32	1.719	.096, ns
least honest	3.21	0.68	1.177	.25, ns
Run 12				
most reliable	4.75	6.12	-0.597	.55, ns
least reliable	-0.08	4.64	-2.213	.03**
most honest	4.46	4.68	-0.080	.94, ns
least honest	6.08	-0.64	2.723	.009***
Run 24				
most reliable	3.92	2.28	0.591	.558
least reliable	1.00	0.64	0.239	.812
most honest	6.38	3.76	0.748	.458
least honest	0.29	2.84	-1.030	.309

** $p<0.05$, ** $p<0.01$

the worst case scenario, or in the best case scenario, or in the most likely scenario.

I chose to carry out an analysis on the representative run (Run 12) that was typical of Cluster, the cluster that produced the best outcomes. I was interested in whether the run reached consensus and, if so, how quickly it did so. I was also interested in whether the dynamics of that consensus development were trivial.

Dynamics in the evolution of a consensus The adoption of the consensus strategy happened relatively quickly, with a sharp reduction in the number of different strategy weightings early in the trading history. Figure 8.11 indicates that a consensus on strategies happened within the first hundred trading rounds, indicating that agents have passed information through the population quickly. While the number of different strategy weighting being used in the drops consistently in each round, this does not indicate that traders are all moving to a single dominant strategy. Figure 8.12 shows the number of agents using each of the different values applied to the weighting on bad deliveries. This shows that the relative popularity of each weighting changes as the overall strategy develops. Strategies may

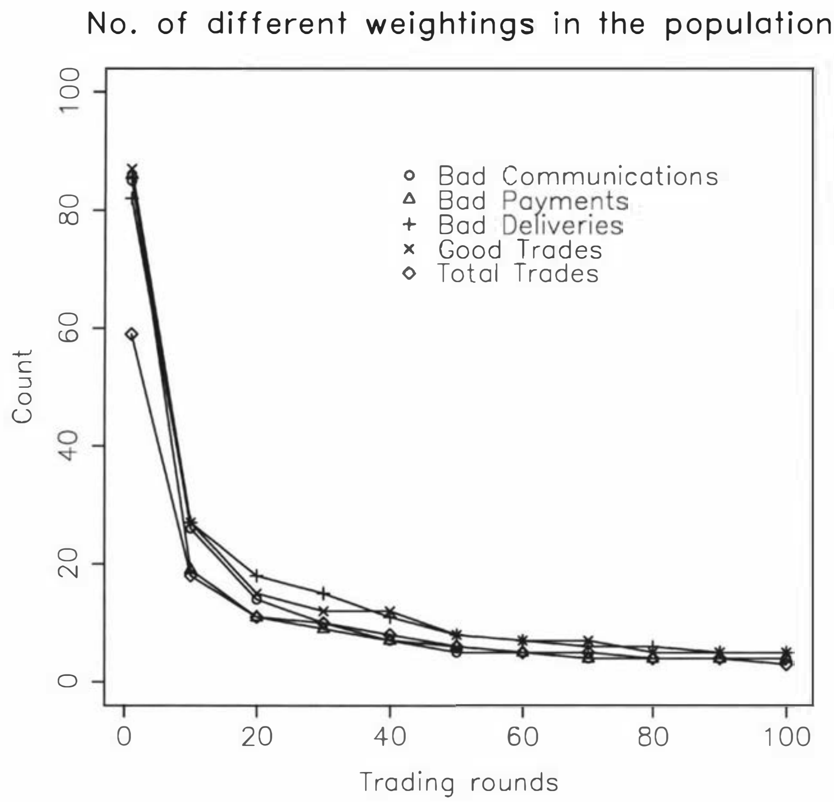

Figure 8.11: Number of different weightings for each element of strategy over the first 100 trading rounds in Run 12

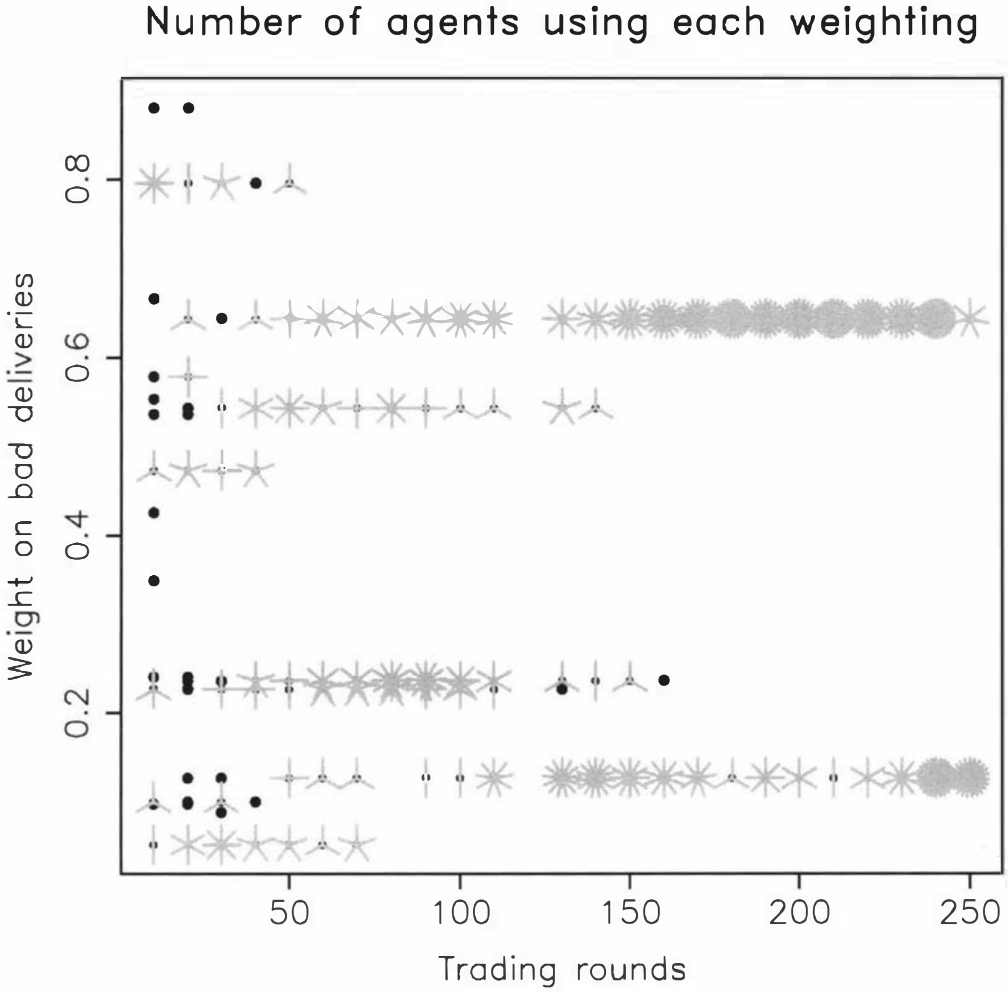

Figure 8.12: Sunflower plot showing adoption of strategy element for weight given to the number of bad deliveries by Agents 1-35 in Run 12.

be relatively popular for a period, but fall out of favour later. Similar plots can be constructed for other strategy elements.

Discussion

All runs of the Trading Model moved toward a consensus on the trust strategies in their populations over the first 250 runs. But while consensus were reached, the strategies that were evolved are not necessarily optimal, and not all were even particularly effective. Rather, their distribution over a number of runs continues to reflect the random uniform distribution of the initial strategy generation.

There were, however, differences in the outcomes produced in different runs. While there was no clear indication as to the ideal number of clusters, grouping the runs into three clusters produces one group in which traders suffer a high number of trade failures due to unreliability, and very low levels of trade failures due to dishonesty, one group in which there are a low number of failures due to unreliability, and a higher level of trade failures due to dishonesty, and one group of runs with moderate levels of failures caused by either unreliability or dishonesty. These groups were most strongly differentiated by the number of trade failures due to unreliability, rather than failures due to dishonesty.

Perhaps counter-intuitively, the overall happiness is lowest in runs that develop the most sensitivity to dishonest traders. It is possible that this sensitivity to dishonest traders, while restricting the number of fraudulent trades, also constrains the number of good trades. The number of good trades is highly correlated with the overall happiness. This is not surprising, because in the model any loss to a fraudster is exactly equalled by the gain made by that fraudster. At a population level, this leaves the determinant of overall happiness being driven by the number of trades that fail due to unreliability: that is, due to a failure to communicate or a failure to make payment. The consequence of this is that, at the population level, a strategy that identifies possibly unreliable traders may be more effective than one that identifies possible fraudsters. The best performing strategies, however, resulted in a drop in number of failures due to unreliability, but constrained the activities of the least honest traders, without restricting lower risk traders.

The results have implications for real-world online auction markets. Where no guidelines are given as to how trading history information might be used, people are left to their own devices in developing their individual strategies. Learning through individual trial and error runs the risk of being a slow and expensive process for the individual, as information is only gleaned by from a traders own personal trading experiences. A form of observational learning allows for the results of many peoples' trading experiences to be incorporated into a trust strategy culture that is specific to that trading environment. Each online market may evolve different strategies, some of which might be more effective than others. There is some evidence that this may be the case, with a recent announcement by eBay that it will no longer provide for auction feedback from sellers. Dishonest or unreliable sellers had been diluting the salience of their own bad trades by generating large volumes of false data through comments on buyers. This problem does not seem to exist on Trade Me.

Possible future development of the model

The Trading Model has scope for further development. In the current model the population is static, with no new entrants, or traders leaving the population. Partly as a result of this, there are no new weightings entering the market that challenge the existing weightings. There is also a lack of experimentation with new weightings among existing traders. Both could be introduced into the model.

The model population is generated in a single batch, which remains fixed for the entire simulation. Real markets, including on-line markets, have scope for entry and exit of traders. This is a particularly active mechanism for dishonest traders in on line trading; having developed a bad reputation under one name, they can leave the market very easily, and enter under an entirely new name, with no reputation. Further development of the model should allow for the exit of participants. This might include those who have unsatisfactory trading experiences, represented by a low Happiness score in this model, as well as those who have developed a bad reputation.

The model is driven to a large extent by the way that traders make their decisions, and the way that they determine how successful they and others have been in trading. In this model, the traders use a simple linear combination of the available data to make a decision as to whether or not to bid on any one item. There is scope for development and experimentation with both the form of the decision-making function, and with the fitness function that is used by the individual traders to determine the success of a strategy.

Further, it is highly unlikely that real world traders use such an explicit process to make their decisions. Rather, descriptions as to how to use the information tend to use fuzzy terms. For example, the eBay site tells traders that "a high Feedback Score and high percentage is a good sign" (eBay, 2006). This recommendation is couched in fuzzy terms: scores and percentages should be "high", and if they are this is "a good sign". There are no indications that help traders translate these fuzzy terms, the meaning of which are likely to be bound to the on-line auction environment. Traders are not told what would be considered a high percentage. For example, while 80% might be considered a high percentage in an exam, it might be regarded as a low percentage of good trades in a trading environment.

Finally, the model uses the quantity *Happiness* as the measure of each agent's success. This quantity is updated at the completion of each trade. In the model, traders judge their own success relative to their partner from their relative accumulated Happiness over all trading rounds. This means that the effects of strategy refinements are likely to take some time to appear in the overall Happiness score, especially in later rounds where a substantial proportion of a total Happiness score might have been assembled before the current strategy had been adopted.

An alternative is to use an incremental measure, rather than a cumulative measure of Happiness. This might be the Happiness at the previous round, or over a group of recent trades. It would not be unrealistic to expect traders to have access to this information, information on the outcome of individual past trades is accessible. But using the incremental measure has implications for the realism of the the model in terms of who information is obtained from. In reality, traders are likely to exchange information within separate networks: social networks of friends, colleagues, and family. These sources of information are outside the on-line trading site, and information on their success is more likely to be in terms of overall trading success, rather than in terms of their last trade. In the current model, the information source is a trader with whom friendly relations might be assumed, following on from an existing trade. The model uses the existing network of traders to double as the social network from whom the strategy information might be obtained. Individual trading rounds are quite brief, and traders do not generally exchange strategy information in the context of a trade. Although information on the last trades might be available, it is unlikely that strategy information would be available within the context of trading.

Each of these possible modifications to the model adds complexity to the model. Doing so raises the need to balance increasing complexity with what is possibly a more realistic agent functionality. Further, while each addition may have the potential to add some element of realism, there is the likelihood that the model develops a realism of form, but at the cost of the insertion of more detail that may be relatively unsupported.

Chapter 9

Conclusion

Constructing the agents

The two agent-based models constructed for this research were designed using quite different approaches to using psychological theory for constructing the agents. The BasicBreaking model was designed to directly use a game theoretic cognitive model of trust game strategies. The cognitive model had itself been generated from a combination of simulation modelling and experimental results. This model was designed at a purely individual level, using only this cognitive model. This produced an entirely abstract model of trust in a particular game situation.

Formal trust games are limited in that they are restricted to particular types of interaction in which trust is engaged. Specifically, they usually target making trust decisions about some externally quantifiable element, such as money in one-on-one interactions. In experimental settings, formal games are usually played between participants that do not know each other, and cannot identify each other. In practice, players often engage in the game-play through a human or computer intermediary. This minimises the information that can be applied to a decision, in many cases restricting the available information to information on the other player's move.

The Trader agent was based on a specific real world situation: how agents might develop ways to identify who to trust in an online auction marketplace from the information available. Again the traders in this model must work with a limited, and domain specific set of information. Rather than the very specific cognitive algorithm of the Basic Breaking model, this used a more generic idea, based on observational learning. In this model, traders can learn elements of a strategy for who to trade with and who to avoid from other traders, and that they would selectively adopt the strategies of successful traders.

Validation

A substantial challenge for agent-based models, and for simulation modelling in general is the question of how to validate the models generated. The very characteristic of large nonlinear systems that at once makes them generate interesting behaviours leads them to generate different results for each run. The result is that a successful model may not produce direct and testable predictions. Rather, a number of runs of the model may produce a range of possible outcomes. Validating the model against data becomes a problem because we usually only have a single "run" in real life. Agent-based modelling can offer a glimpse of the range of possible results, and it can give an indication of the relative likeliness of these possible results. What it cannot do, except in limited and not especially interesting models, is predict which possible result will obtain on any one occasion.

The Basic Breaking model demonstrated this to some extent. It identified pairings that are likely to be most compatible, but the stability of behaviour generated when all of the agents are using one of a limited range of strategies that are known to be successful produced a model that is very stable, and did not behave in any interesting ways.

One means proposed for validation of simulation models is through the validation of the bases on which the model was constructed. If we accept that this would, at least, help in the construction of valid models, this leaves us in the position of needing solid and validated theories to draw on to build the model. This is the domain of social psychology; the development and testing of solid and validated theory of individual behaviour in social settings. The challenge for the agent-based modeller is how to translate these theories into a form that is usable in the model. One suggestion is that the theory should be translated into a formal logic form. This is a controversial idea among agent-based modellers. Whether or not the theory is translated into a formal logic form, there is a degree of formalisation needed to write the software for the model.

An effect of this need for the at least minimal formalisation of the theory is that the very act of constructing an agent-based model is sufficient to force the modeller to be very specific about theory. Gigerenzer (2000) has suggested that scientific theory is driven by the tools that scientists use.

This is apparent when attempting to develop an agent-based model using existing theory from social psychology: the modeller needs theory that can be translated to a very specific statement. In the example of the Trader model, observational learning provided a theoretical base that could readily be translated into a model format. Observational learning can be modelled if we can identify what is to be learned, and a means that the thing to be learned can be observed, and combine this with some sort of memory device.

That said, in the context of modelling online trading, I needed to be specific about what was to be learned by the agents. In this case, what is to be learned is the strategy by which an agent decides whether or not to trade with another agent. This places a much more domain specific requirement for theory, that is not available in the trust literature. Further, programming agent behaviours for a population of agents means that we need know not only how the characteristic is expressed on average, but also the range of possible expressions that we might expect to find in the population. This results in a need to have more understanding of how people differ, rather than the location of the average. The Basic Breaking model, in particular suffered from the population being too homogeneous, and the lack of cheaters meant that the agents were never challenged to the point where they were triggered to break relationships, even with very large deviations in the exact values used as decision-making parameters. Agent-based models of trust depend on being able to represent the range of deviant behaviours that might be encountered.

This also applied with respect to the Trader model, which needed some distribution of dishonest and unreliable traders. Basically, models of trust need cheaters, and so we need to understand cheaters' behaviours as well as we might understand successful honest traders' behaviours. While the study identified the most popular strategies, and was able to associate these with successful strategies, it provided no information on the diversity of strategies. The results generated were in terms of identifying common strategy clusterings, but modelling on the basis of the findings needed a fuller description of the strategy clusters to be able to put the information into an agent-based model. Without this, the behaviour of the Basic Breaking model was dull, because all of the agents were all cooperative.

This need for information on diversity also has a positive side for agent-based models, particularly in relation to their application. While they pose a challenge in terms of the theoretical data required, agent-based modelling gives us something to do with diversity information. Each agent in a model can be equipped with an individual set of characteristics.

While the main thrust of Gigerenzer's (2000) argument about tools to theories heuristics is that they can restrict the development of the theory, this is not the only possibility. Within the field of artificial intelligence the unidirectional nature of theoretical influence has been less clear cut. While we may draw theoretical ideas from the devices that we are using, ideas from psychology are also adopted for use in computing devices, particularly software devices. It might be more balanced to see the development of new ideas in artificial intelligence and psychology as inspiring different ways of thinking in both disciplines. In fact, we might go further and say that developing an agent-based model brings a level of specificity that forces a different way of thinking about how individuals are characterised, how they process information, and how they interact with other individuals.

The form of the results from an agent-based model

The output from an agent-based model tends to be in a series of numerical data. This can be presented either as a graphical display from within the ABM programme, or as an output data table sets can then be analysed using external tools. These data sets can be very large. For example, the data set generated by the Trader model over a number of runs produces a 2.6MB data set.

The modeller's task at the analysis stage is to identify patterns in these large data sets. Statistical techniques provide useful tools for doing this, but their interpretation in this use is a little different to the usual interpretation in psychology. Hypothesis testing is the dominant analysis used in quantitative social psychology. But the logic of hypothesis testing is that it is interpreted as the likelihood that a pattern could have arisen by chance in a sample, the alternative being that it reflects a pattern existing in the population. The meaning of these statistical analyses is less clear when the sampling is of one of a theoretically infinite number of possible runs of the model. In a simulation model, it is less clear what is sampled, from what population. Is a run a single sample of all possible runs? Or a sample of, say 100 agents from an infinite population of possible agents?

Statistical analyses do remain as a solid and well-founded form of analysis, but we need to consider the logic that we call on to use these in a simulation modelling mode. Here we need to find patterns generated by the model, and to do so we might make use of exploratory and descriptive statistical techniques, rather than hypothesis testing techniques.

But there is a further issue with using statistical methods. Returning to the fundamental mathematics behind statistical modelling, inferential analysis involves proposing a form of mathematical model that may represent the process underlying the data. That model includes elements representing systematic relationships with variables, and an element representing a random, non-systematic, noise component. For example, multiple linear regression is used to fit a typical linear equation of the form $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$, in which y is a dependent variable, x_1, x_2, and x_3 are independent variables, and ϵ is a random noise component. Inference is the process of producing estimates parameters from available data, to extract patterns from noise. In this case, patterns are described by the overall form of the equation, including derived estimates for β_1, β_2, and β_3. Extracting a pattern from noise presupposes a particular form for the noise component. Conventional research provides a number of potential sources for this noise: for example it might come from other internal influences on the people being researched, from measurement instruments, or from the environment. In contrast, in the case of an agent-based model, the only source of noise is from within the model itself. Statistical analysis in this context cannot be used to eliminate noise that is external to the process. Rather it is used to detect patterns within more complex overall result.

In common with the task of pattern detection in artificial intelligence more generally, it may still be difficult to pick up patterns. The use of other data visualisation techniques may be helpful. Agent-based modelling programmes are complete with a set of visualisation tools; for example, Repast can generate Quicktime (.mov) format movies from iterated graphical output, but these do not necessarily translate meaningfully into a paper reporting medium. Further, emergence, and the data in general may not be suitable for the same form of analysis as patterns develop. The Trading Model provides an example of this. It starts with what is, effectively, a set of continuous numerical data. After a number of turns, the number of strategies is reduced as individuals adopt the strategies of the most successful traders, shifting the form of the data to a more non-parametric form as the count of individuals adopting each strategy grows.

Developing a model, running it, and extracting a result

The objective of this work has been to develop an agent-based model of trust within a social network. Development of such a model involves a number of activities. One of the first activities is to identify the types of phenomena that agent-based modelling is best suited to. There are two key characteristics that signal the need for some form of dynamic analysis, and in turn simulation to access the dynamics. One of these characteristics is the presence of dynamic, time located effects, or development of a phenomenon in time. The second characteristic is that the phenomenon is located in a system involving the interaction of a number of individuals.

While either or both of these characteristics might be sufficient to suggest that a form of analysis that can address a nonlinear dynamic system, there are a number of ways that this might be carried out. It is possible, in some instances, to collect time series data directly. If it is not possible to collect data directly, there are two options. The problem might be reformulated to attempt to avoid the dynamic and aggregation effects. For example, we might wait until a system reaches equilibrium before making measurements. The difficulty with this approach is knowing when it has done so. Finally, we might use simulation modelling to generate dynamic data. Agent-based modelling is one of a number of techniques that might be used to generate data.

The choice of method will depend on the characteristics of the phenomenon of interest, and on the mechanisms that might lie behind these. Agent-based modelling is a candidate where theory suggests that the interacting behaviours of individuals might be important, and when we have theory by which we might formulate the behaviours of those individuals.

Characteristics indicating that a dynamic approach should be considered

Dynamics

Agent-based models provide a means to investigate some features of dynamic behaviour, specifically the complex behaviours that emerge from large systems of nonlinear elements. Time is an implicit part of many concepts in social psychology, and may come into play at any time that there is change. The likelihood that dynamics come into play is raised in systems of a number of individuals. Despite this, time is largely an absent dimension in research and theory. In part, this is because when applying statistical analyses to real data we meet the constraint that it is rare that we can generate sufficient data points to meet the assumptions of time series analysis. A practical compromise is to reduce the measurement intervals to single time period, or a very few intervals, but this tends to make time invisible.

One effect of the relatively low profile of time is that social psychology has a strange relationship with time as a construct, especially if we compare social psychology with developmental psychology, in which time is an integral part of theorising. In comparison, in social psychology explicit theories about the time dependence of phenomena are limited, and propose that the linear conception of time may not always be appropriate in social psychology.

While agent-based models do allow us to explore patterns of phenomena that might develop, in general the iteration steps in an agent-based model do not map onto a time scale unless the model has been designed explicitly to do so. Although they are not explicitly scaled, agent-based models do allow us to explore some aspects of the dynamics of systems of individuals. System dynamics exist, at least implicitly, along a time dimension. This has been raised as a weakness of dynamic methods in general. On the other hand, a lack of time scaling may not necessarily be a disadvantage in a social psychology setting. It has been suggested that thinking in terms of continuous linear time are not necessarily the most appropriate in social psychology, where continuous time is applied to discontinuous processes. Agent-based models are often based on a turn-taking basis, a computational necessity where the model is located in a single serial computation process. There are alternatives, notably the ability to use modelling tools like NetLogo on a network of computers, or running tools like Repast on computer clusters, but these are hardware intensive.

Agent-based models allow us to investigate the stability of systems, whether they find a stable position, and what that stable position might be. Both the Basic Breaking and the Trader models generated stable positions. While the Basic Breaking model did not generate any interesting behaviour, the Trader model produced a stable position on every run. The stability of the model was a consequence of the design of the model, as information was passed in one direction only, from more successful to less successful traders.

While each run produced a stable position, there was no particular pattern in values of weightings in these stable positions. Although strong attractors developed for each run, as evidenced by the stable outcomes generated, the values of the weightings was not a factor in locating these attractors.

As discussed in Chapter 2, nonlinear dynamic systems can produce a variety of attractors, including stable point attractors, limit cycles, attractors with a number of periodic components, and strange attractors. Neither of the models produced other than single point stable attractors. These stable outcome positions from the single point attractors in the Trader model were not derivable from the starting characteristics of the individual agents.

These stable point attractor outcomes are less dramatic than some of the more exotic phenomena that nonlinear dynamic systems can produce. The literature on dynamic systems in social psychology has tended to concentrate on demonstrating that the entire range of phenomena that can be generated by nonlinear systems are also tractable to analysis, and thus amenable to application in research. While it is valuable to have an accessible survey of methods that can be used to access dynamic behaviours, these can seem somewhat daunting.

One side-effect that concentrating on some of the more complex outcomes is that they can make less dramatic, but still important, phenomena seem less exciting. This is a shame, because understanding the stability of social systems is important to real world outcomes. Understanding the existence of system states that are stable, and the robustness of that stability under perturbation is key to understanding the risks, benefits, and possible outcomes that might stem from an intervention. In the setting of the Trader model, for example, it appears that if people simply adopt the strategies of more successful trader friends, the resulting strategy may not be optimal. Leaving this as the means by which people devise these strategies may mean that less than optimal strategies become entrenched, a version of magical thinking at a systems level.

Finally, investigation of dynamic systems in general, and particularly the investigation of system behaviour in the presence of the more complex attractors depends on a large volume of data; the more complex the dynamical behaviour, the more data are needed to analyse the behaviour. These methods encounter a problem in investigating time-located phenomena in the social sciences: it is very difficult to collect sufficient real-world data to investigate these directly. Simulation is one method that allows us to generate enough data for analysis.

Aggregation

The very nature of agent-based models, as larger models built from a number of smaller models of individuals, places them at the boundary between between the disciplines of sociology and social psychology. Being at this boundary, one of the criticisms that agent-based modelling encounters is that modelling larger entities at the level of the individual is flawed, as it fails to take into account for the effects of structures.

One of the major claims made by the proponents for analysis at the level of structure is that structure is an emergent. In this context, emergent is taken to mean that it cannot be understood by being reduced to its parts, that it cannot be explained in terms of the properties of its parts, and that emergent properties exhibit downward causation. The definition of emergence in terms of irreducibility is understandable as long as there is no mechanism that might explain how low level features can combine to produce emergents.

Arguments that features of social structure cannot be explained other than in their own terms need to be reviewed in the light of the greater understandings of complex systems. Relatively recent work in the mathematics of complex and chaotic systems has demonstrated that systems can produce a rich variety of outcomes, from simple stable positions, through simple oscillatory patterns, to immensely complex patterns. While some outcomes may be straightforward, others are not obvious combinations of the effects of a number of elements. While it is not entirely clear that these phenomena lie behind the development of elements of social structure, it is also far from clear that they do not.

The very nature of agent-based modelling encounters these issues about whether the individual or the social structure are the more appropriate level of analysis. In part, these might be seen as a specific instance of a more general issue as to whether analysis at a component or a systems level is more appropriate. The terminology used in agent-based modelling tends to lead towards us thinking of agents in terms of individuals, but there is no particular reason to suggest that agent-based modelling should be restricted to modelling human individuals. For example, there may be situations where it is not the individual that is of particular interest, rather it is the relationship between individuals that is of interest. This is a more explicitly social network oriented element, and more naturally leads to thinking in terms of systems than does thinking about individuals and the nature of their possible and significant interactions.

As with the dynamics of social systems, the exploration of potential mechanisms lying behind social structures is very difficult to carry out directly on real world systems. Again, simulation provides a means to explore of these possible mechanisms.

Thinking about social psychology

One of the potential strengths of agent-based modelling is that it demands particular ways of thinking about individuals, relationships, and how things develop in time. Firstly, even at the design stages, it forces us to think in terms of systems. This is not necessarily a natural way to think about social processes, or about systems processes more generally. Moving into systems thinking, specifically to systems that contain nonlinear elements, has the effect that we cannot ignore the effect that these systems have on our ability to make predictions. This, in turn, added to the nature of interaction with an agent-based model has an effect on how we interact with theory.

Systems thinking

Despite requiring specific thinking and theorising at the individual level for the construction of the model, building an agent-based model also requires thinking in terms of systems, both in the selection of individual characteristics, actions and communications to be represented, and in the work in experimenting with the model. This was a factor in making the Basic Breaking model dull; the algorithms applied were drawn in terms of individual behaviours, and interactions that are contained within a dyad. This model was deficient both in terms of the individuals and dyads located in a network of other players, and in the more specific criteria for generating swarm behaviours (Bonabeau et al., 1999), particularly in the absence of any mechanism for positive feedback.

Predictions

One of the more powerful ideas in the philosophy of science is the idea that science progresses by making predictions and by testing these. Thus one of the essential components of planning conventional quantitative research methods is that the researcher identifies hypotheses, predictions that can be tested, and questions based on these. This form of research question depends on being able to make predictions that we might reasonably expect to be able to test. When working with systems that may generate emergent features, and in which outcomes are so dependent on initial conditions, making predictions becomes more problematic. Research questions take on a different character, more along the lines of "What are the potential outcomes when these conditions apply?" The detailed questions that might guide searches for patterns are not necessarily obvious before modelling is carried out. The results of a simulation run, and the thinking about the results and how the system has generated them gives rise to new questions and experiments. This interaction with the system, and results is a process of theory development.

Theory

The active interaction with the model, its behaviour, and the incorporation of the understanding of these in the development of the model means that engaging with an agent-based modelling serves to make this method both theory intensive and theory generative. Further, it adds a different form of test for existing theory in social psychology. Attempting to apply theory in an agent-based model tests our understanding of the theory, by challenging us to work backwards and build something from theoretical elements. It also tests whether the theory is open, general, and complete enough to be applied and used generatively: can this theory be applied to an individual, with meaning, and with effect? This applies both for the simple cognitive algorithms, such as those that were used to construct the Basic Breaking model, and for bigger theories, such as Social Cognitive Theory.

While agent-based modelling is not a universal research tool, in particular situations it offers the possibility of accessing phenomena that are not otherwise accessible, and for developing understanding in systems contexts. This means that agent-based modelling has the potential to be useful in social psychology, both in the development of theory, and in providing an opportunity to interact with that theory.

Summary

This thesis has reported the development of two agent-based models of trust as it operates within networks of individuals. Understanding trust in this setting is important, as it enables the formation and maintenance of social cooperation. Doing so involves finding a way to address the aggregation and of individual behaviours, where the behaviours are influenced by the behaviours of other people in the network. This may result in systems that generate complex dynamic behaviour that are difficult to address using survey and experimental methods, and conventional statistics.

One way that such systems it may be approached is through computer simulation using agent-based models. This thesis describes the development of two agent-based models of trust. Agent-based modelling is a novel method within the discipline of social psychology. The thesis first describes what agent-based modelling is, describes some of the situations in which it might be applicable, discusses how it might apply to modelling individuals in a social setting, and discusses the experience of developing the model.

The first model was based on a theoretical cognitive model of behaviour within a formal game that has been claimed to involve trust, the Investor Game. This model showed that a population in which all individuals are pursuing similar optimal strategies does not generate any of the interesting behaviours that we would expect to see in real-world interactions involving trust and cooperation. This tends to suggest that modelling trust behaviours also requires modelling behaviours that are untrustworthy, and representing a full range of potential behaviours, including outliers.

The second model was based on a more naturalistic setting, on-line peer-to-peer trading through sites such as New Zealand's Trade Me, or eBay. In this model, individual traders, represented by agents, carry characteristics that determine their reliability and honesty, and attempt to find effective strategies for identifying other traders' trustworthiness. They exchange information, in a form of social learning, as they attempt to find an optimal trust strategy. This model suggests that, while providing traders with minimal guidance on strategies and allowing them to search for the best strategies may result in them finding effective strategies, this is not the only possible outcome. Somewhat surprisingly, effective trust strategies acted to contain unreliability, rather than dishonesty.

My experience of using agent-based modelling for this research was that this is a viable and useful method with potential for wider application in

social psychology. It is particularly well suited to providing ways that we might apply and explore some of the important theoretical ideas in social psychology, such as Bandura's ideas about triadic causality, that are otherwise difficult to address through conventional means. Developing the model involves an explicit interaction with theory, as programming the model is theory made concrete.

References

Abreu, D., & Sethi, R. (2003). Evolutionary stability in a reputational model of bargaining. *Games and Economic Behavior*, *44*, 195–216.

Ajzen, I. (1985). Action-control: From cognition to behavior. In J. Kuhl & J. Beckmann (Eds.), (pp. 11–39). Heidelberg: Springer.

Allen, R. E. E. (1980). *The concise Oxford dictionary of current English*. Oxford: Clarendon Press.

Archer, M. S. (2000). *Being human: The problem of agency*. Cambridge: Cambridge: Cambridge University Press.

Axelrod, R. (1997a). Advancing the art of simulation in the social sciences. *Complexity*, *3*(2), 16–22.

Axelrod, R. (1997b). *The complexity of cooperation: Agent-based models of competition and collaboration*. Princeton, New Jersey: Princeton, NJ: Princeton University Press.

Baier, A. C. (1986). Trust and antitrust. *Ethics*, *96*, 231–260.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1999). Social cognitive theory of personality. In O. P. Pervin L. A.; John (Ed.), *Handbook of personality theory* (pp. 154–196). New York: Guilford Press.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annual Review of Psychology*, *52*, 1–26.

Bertalanffy, L. v. (1971). *General system theory*. London: Allen Lane.

Bierhoff, H. W. (1992). Trust and trustworthiness. In L. Montada, S.-H. Filipp, & M. J. Lerner (Eds.), *Life crises and experiences of loss in adulthood*. New Jersey: Lawrence Erlbaum.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of the Sciences*, *99*(Suppl. 3), 7280–7287.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. Oxford: Oxford University Press.

Bosse, T., & Treur, J. (2006). Formal interpretation of a multi-agent society as a single agent. *Journal of Artificial Societies and Social Simulation*, *9*(2), Retrieved June 19, 2006, from http://jasss.soc.surrey.ac/9/2/6.html.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology*. Cambridge, MA: MIT Press.

Bratman, M. E. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Camerer, C. F. (2003a). *Behavioral game theory: Experiments in strategic interaction.* Princeton: Russell Sage Foundation and Princeton University Press.

Camerer, C. F. (2003b). Behavioural studies of strategic thinking in games. *Trends in Cognitive Sciences*, *7*(5), 225–231.

Castelfranchi, C., & Falcone, R. (2000). Trust and control: a dialectic link. *Applied Artificial Intelligence*, *14*, 799–823.

Checkland, P. (1999). *Systems thinking, systems practice : includes a 30-year retrospective.* Chichester, New York: John Wiley.

Chown, E., Jones, R. M., & Henninger, A. E. (2002). An architecture for emotional decision-making agents. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1.* New York: ACM Press.

Cohen, J., & Stewart, I. (1995). *The collapse of chaos: Discovering simplicity in a complex world.* London: Penguin.

Colman, A. M. (2003). Cooperation, psychological game theory, and limits of rationality in social interaction. *Behavioral and Brain Sciences*, *26*, 139–198.

Conte, R., Gilbert, N., & Sichman, J. S. (1998). MAS and social simulation, a suitable committment. In J. G. Carbonell, J. S. E. Siekmann, J. S. Sichman, R. Conte, & N. V. E. Gilbert (Eds.), *Multi-agent systems and agent-based simulations.* Berlin: Springer.

Cook, K. S., & Cooper, R. M. (2003). Experimental studies of cooperation, trust, and social exchange. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 209–244). New York: Russell Sage Foundation.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Cosmides, L., & Tooby, J. (1997). *Evolutionary psychology: a primer.* Retrieved 29 September, 2004, from http://www.psych.ucsb.edu/research/cep/primer.html.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: an experimental study. *Econometrica*, *69*(5), 1193–1235.

Couch, L. L., Adams, J. M., & Jones, W. H. (1996). The assessment of trust orientation. *Journal of Personality Assessment*, *67*(2), 305–323.

Couch, L. L., & Jones, W. H. (1997). Measuring levels of trust. *Journal of Research in Personality*, *31*(3), 319–336.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*, 260–281.

Crimes Act. (1961a). Statutes of New Zealand, s.193.

Crimes Act. (1961b). Statutes of New Zealand, s.196.

Damasio, A. R. (1994). *Descartes error: Emotion, reason, and the human brain.* New York: Putnam.

Deutsch, M. (1958). Trust and suspicion. *Journal of conflict resolution*, *2*, 265–279.

Doran, J. E. (2005). *Iruba: An agent-based model of the guerrilla war process.* Retrieved 11 Aug, 2007, from http://cswww.essex.ac.uk/staff/doran/34.PDF.

Doran, J. E. (2006). Agent design for agent based modelling. In F. C. Billari, T. Fent, A. Prskawetz, & J. Scheffran (Eds.), (pp. 215–223). Heidelburg: Physica-Verlag (Springer).

eBay. (2006). *What is feedback?* Retrieved 11 December, 2006, from http://pages.ebay.com/help/feedback/questions/feedback.html.

Edmonds, B. (1998). *Social embeddedness and agent development.* Retrieved 27 September, 2004, from http://cfpm.org/cpmrep46.html.

Edmonds, B. (2002). Reasoning about rational agents. *Journal of Artificial Societies and Social Simulation*, *5*(1), Retrieved, 20 Aug 2005, from http://jasss.soc.surrey.ac.uk/5/1/reviews/edmonds.html.

Edmonds, B., & Moss, S. (2005). From kiss to kids an anti-simplistic modelling approach. In P. Davidsson et al. (Eds.): (Ed.), *Multi Agent Based Simulation 2004* (Vol. 3415, pp. 130–144). Berlin: Springer.

Elder-Vass, D. (2005). Emergence and the realist account of cause. *Journal of Critical Realism*, *4*(2), 315–338.

Elder-Vass, D. (2007). For emergence: Refining Archer's account of social structure. *Journal for the Theory of Social Behaviour*, *37*(1), 25–44.

Enquist, M., Arak, A., Ghirlanda, S., & Wachtmeister, C. A. (2002). Spectacular phenomena and limits to rationality in genetic and cultural evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *357*, 1585–1594.

Evans, J. S. B. T. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Fodor, J. A. (2001). *The mind doesnt work that way: the scope and limits of computational psychology.* Cambridge, MA: MIT Press.

Frigg, R., & Hartmann, S. (2006). Models in science. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy* (pp. Retrieved January 3, 2008, from http://plato.stanford.edu/entries/models-science/TwoExtSynSemVieThe).

Fromm, J. (2002). *On emotion controlled architectures for intelligent autonomous agents.*

Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity.* New York: Free Press.

Gaylord, R. J., & Nishidate, K. (1996). *Modeling nature: Cellular automata simulations with Mathematica.* New York: Springer-Verlag.

Geert, P. van. (1997). Time and theory in social psychology. *Psychological Inquiry*, *2*(2), 143–151.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (1999). The belief-desire-intention model of agency. In J. G. Carbonell, J. S. E. Siekmann, J. P. Mülle, M. P. Singh, & A. S. V. E. Rao (Eds.), *Intelligent agents V: Agent theories, architectures, and languages.* Berlin: Springer-Verlag.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world.*

Oxford: Oxford University Press.

Gigerenzer, G., Todd, P., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Gilbert, N., & Bankes, S. (2002). Platforms and methods for agent-based modeling. *Proceedings of the National Academy of the Sciences*, *99*(Suppl. 3), 7197–7198.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the social scientist*. Buckingham: Open University Press.

Gleick, J. (1988). *Chaos: Making a new science*. London: Sphere Books.

Glennan, S. (2000). A model of models. *Retrieved from http://philsci-archive.pitt.edu/archive/00001134/00/Models1.4.pdf on 20 Aug 2005*.

Gooding, D. C., & Addis, T. R. (2006). Modelling experiments as mediating models. In L. Magnini (Ed.), *MBR'04: Model-based reasoning in science and engineering: Proceedings, Fourth International Conference on Model-Based Reasoning*.

Goto, S. (1996). To trust or not to trust: the situational and dispositional determinants. *Social behavior and personality*, *24*(2), 119–132.

GSS 2000 Codebook. (2000). *Retrieved May 5, 2004 from http://www.webuse.umd.edu/sdaweb/gss2000/Doc/gss20015.htm*.

Gurtman, M. B. (1992). Trust, distrust and interpersonal problems: a circumplex analysis. *Journal of Personality and Social Psychology*, *62*(6), 989–1002.

Haag, D., & Kaupenjohann, M. (2001). Parameters, prediction, post-normal science and the precautionary principle - a roadmap for modelling for decision-making. *Ecological Modelling*, *144*, 45–60.

Hahn, C., Fley, B., Florian, M., Spresny, D., & Fischer, K. (2007). Social reputation: a mechanism for flexible self-regulation of multiagent systems. *Journal of Artificial Societies and Social Simulation*, *10*(1), Retrieved 11 August, 2007, from http://jasss.soc.surrey.ac.uk/10/1/2.html.

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, *10*(4), 371–388.

Hales, D. (1998). Artificial societies, theory building and memetics. *Retrieved February 15, 2005, from http://citeseer.ist.psu.edu/117691.html*.

Halpin, B. (1999). Simulation in sociology. *American Behavioral Scientist*, *42*(10), 1488–1508.

Hardin, R. (2002). *Trust and trustworthiness*. New York: Russell Sage Foundation.

Harris, I. (2003). The Sims online. *PC Games Addict Tips Special*, *2*, 8–16.

Hastie, R., & Stasser, G. (2000). Computer simulation methods in social psychology. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 85–114). Cambridge, UK: Cambridge University Press.

Hegselmann, R. (1998). Modeling social dynamics by cellular automata. In B. G. Liebrand, A. Nowak, & R. Hegselmann (Eds.), *Computer modeling of social processes*. London: Sage Publications.

Hegselmann, R., & Flache, A. (1998). Understanding complex social dynamics: A plea for cellular automata based modelling. *Journal of Artificial Societies and Social Simulation*, *1*(3), Retrieved November 11, 2006, from http://www.soc.surrey.ac.uk/JASSS/1/3/1.html.

Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*, 487–490.

Henninger, A. E., Jones, R. M., & Chown, E. (2003). Behaviors that emerge from emotion and cognition: Implementation and evaluation of a symbolic-connectionist architecture. In *Proceedings of the second international joint conference on autonomous agents and multiagent systems: part 1.* New York, NY: ACM Press.

Holland, J. H. (1998). *Emergence: from chaos to order*. Oxford: Oxford University Press.

Hollis, M. (1998). *Trust within reason*. Cambridge: Cambridge University Press.

Holton, R. (1994). Deciding to trust, coming to believe. *Australasian journal of philosophy*, *72*(1), 63–76.

Hout, M. (2003). *Money and morale: What growing inequality Is doing to Americans' views of themselves and others.* Retrieved on 31 July, 2007 from http://www.russellsage.org/publications/workingpapers.

Hsee, C. K., & Abelson, R. P. (1991). Velocity relation: Satisfaction as a function of the first derivative of outcome over time. *Journal of Personality and Social Psychology*, *60*(3), 341–347.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Inglehart, R. (1997). *Modernization and postmodernization: cultural, economic, and political change in 43 societies*. Princeton: Princeton University Press.

Jaffe, K., & Cipriani, R. (2007). Culture outsmarts nature in the evolution of cooperation. *Journal of Artificial Societies and Social Simulation*, *10*(1), Retrieved 11 August, 2007, from http://jasss.soc.surrey.ac.uk/10/1/7.html.

Jennings, N. R., Sycara, K., & Woolridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, *1*, 275–306.

Jones, K. (1996). Trust as an affective attitude. *Ethics*, *107*(1), 4–25.

Kahneman, D. (2003). A perspective on judgement and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kashima, Y., McKintyre, A., & Clifford, P. (1998). The category of the mind: Folk psychology of belief, desire, and intention. *Asian Journal of Social Psychology*, *1*(3), 289–313.

Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, *95*, 3–36. (mid-point between extreme reductionism and explicit dualisms. Novel and unpredictable , not explainable or mechanistically reducible. And downwards causation)

King, A. (1999). Against structure: A critique of morphogenetic social theory. *Sociological Review*, *47*, 199–217.

King, A. (2007). Why I am not an individualist. *Journal for the theory of social behaviour*, *37*(2), 211–219.

Knack, S., & Zak, P. J. (2002). Building trust: public policy, interpersonal trust, and economic development. *Supreme Court Economic Review*, *Annual 2002*, 91–107.

Krosnick, J. A., Anand, S. N., & Hartl, S. P. (2003). Psychosocial predictors of heavy television viewing among preadolescents and adolescents. *Basic and Applied Social Psychology*, *25*(2), 87–110.

Kukla, A., & Walmsley, J. (2006). *Mind: A historical and philosophical introduction to the major theories.* Indianapolis: Hackett Publishing Company.

Lahno, B. (2001). On the emotional character of trust. *Ethical theory and moral practice*, *4*(2), 171–189.

Lansing, J. S. (2002). "Artificial societies" and the social sciences. *Artificial Life*, *8*(3), 279–292.

Lansing, J. S., & Miller, J. H. (2003). *Cooperation in Balinese rice farming.* Retrieved on June 23, 2006, from http://www.santafe.edu/research/publications/workingpapers/03-05-030.pdf.

Latané, B., & Nowak, A. (1994). Attitudes as catastrophes: From dimensions to categories with increasing involvement. In R. R. Vallacher & A. Nowak (Eds.), *Dynamical systems in social psychology.* San Diego: Academic Press.

Lawrence, J. W., Carver, C. S., & Scheier, M. F. (2002). Velocity towards goal attainment in immediate experience as a determinant of affect. *Journal of Applied Social Psychology*, *32*(4), 788–802.

Lehmann, J. M. (1993). *Deconstructing Durkheim: a post-structuralist critique.* London: Routledge.

Levine, M. (2003). Times, theories and practices in social psychology. *Theory and psychology*, *13*(1), 53–72.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, *23*(3), 438–458.

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985.

Luger, G. F. (2002). *Artificial intelligence: structures and strategies for complex problem solving* (4th ed.). Harlow, England: Addison-Wesley.

Luke, S., Cioffi-Revilla, C., Panait, L., & Sullivan, K. (2004). *MASON: A new multi-agent simulation toolkit.* Retrieved 4 May, 2007, from http://cs.gmu.edu/ eclab/projects/mason/publications/Agent2003.pdf.

MacDonald, A. P. J., Kessel, V. S., & Fuller, J. B. (1972). Self-disclosure and two kinds of trust. *Psychological Reports*, *30*, 143–148.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*, 101–121.

McCabe, K. A., Smith, V. L., & LePore, M. (2000). Intentionality detection and "mindreading": Why does game form matter? *Proceedings of the National Academy of the Sciences*, *97*(8), 4404–4409.

McFadzean, D. (1994). *SimBioSys: A class framework for biological simulations.* Retrieved on May 30, 2006, from http://www.lucifer.com/ david/thesis.

McKenzie, C. R. M. (2003). Rational models as theories - not standards - of behavior. *Trends in cognitive sciences*, *7*(9), 403–406.

Miller, A. S., & Mitamura, T. (2003). Are surveys on trust trustworthy? *Social Psychology Quarterly*, *66*(1), 62–70.

Minar, N., Burkhart, R., Langton, C., & Askenazi, M. (1996). *The Swarm simulation system: a tool for building multiagent systems.* Retrieved on May 30, 2006, from http://www.swarm.org/wiki/Image:MinarEtAl96.pdf.

Morcol, G. (2001). What is complexity science: Postmodernist or postpositivist. *Emergence*, *3*(1), 104–119.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), (pp. 10–37). Cambridge: Cambridge University Press.

Mosler, H.-J. (2006). Better be convincing or better be stylish? a theory based multi-agent simulation to explain minority influence in groups via arguments or via peripheral cues. *Journal of Artificial Societies and Social Simulation*, *9*(3), Retrieved 11 August, 2007, from http://jasss.soc.surrey.ac.uk/9/3/4.html.

Mosler, H.-J., Schwarz, K., Ammann, F., & Gutscher, H. (2001). Computer simulation as a method of further developing a theory: Simulating the elaboration likelihood model. *Personality and Social Psychology Review*, *5*(3), 201–215.

Moss, S., & Edmonds, B. (2005). Towards good social science. *Journal of Artificial Societies and Social Simulation*, *8*(4), Retrieved 11 August, 2007, from http://jasss.soc.surrey.ac.uk/8/4/13.html.

Moss, S., Gaylard, H., Wallis, S., & Edmonds, B. (1998). SDML: A multiagent language for organizational modelling. *Computational and mathematical organization theory*, *4*(1), 43–69.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*(2), 77–101.

Newman, D. V. (1996). Emergence and strange attractors. *Philosophy of Science*, *63*(2), 245–261.

North, M. J., Collier, N. T., & Vos, J. R. (2006). Experiences creating three implementations of the Repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation*, *16*(1), 1–25.

Nowak, A., & Lewenstein, M. (1994). Dynamical systems in social psychology. In R. R. Vallacher & A. Nowak (Eds.), (pp. 17–53). San Diego: Academic Press.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: a dynamic theory of social impact. *Psychological Review*, *97*(3), 362–376.

Omodei, M. M., & McLennan, J. (2000). Conceptualizing and measuring global interpersonal mistrust-trust. *Journal of Social Psychology, 140*(3), 279–294.

O'Sullivan, D., & Haklay, M. (2000). Agent-based models and individualism: Is the world agent-based? *Environment and Planning A, 32*(8), 1409–1425.

Parker, M. T. (2000). *Ascape: abstracting complexity.* Retrieved on May 30, 2006, from http://www.brook.edu/es/dynamics/models/ascape.AbstractComplex.doc.

Pinker, S. (1998). *How the mind works.* London: Penguin.

Power, H. M., & Simpson, R. J. (1978). *Introduction to dynamics and control.* London: McGraw-Hill.

Puddifoot, J. E. (2000). Some problems and possibilities in the study of dynamical social processes. *Journal for the theory of social behaviour, 30*(1), 79–97.

Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in America. *PS: Political Science and Politics, 28*, 664–683.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation, 82*(9), 609–623.

Rao, A. S. (1996). AgentSpeak(L): BDI agents speak out in a logical computable language. In R. van Hoe (Ed.), *Seventh European workshop on modelling autonomous agents in a multi-agent world.* Eindhoven, The Netherlands.

Ray, T., & Liew, K. M. (2003). Society and civilization: An optimization algorithm based on the simulation of social behavior. *IEEE Transactions on Evolutionary Computation*, 7(4), 386–396.

Reeves-Ellington, R. (2004). Trustscapes and distructscapes: A multilevel approach for understanding stability and change. *Multi-Level Issues in Organizational Behavior and Processes Research in Multi-Level Issues, 3*, 91–148.

Resnick, M. (1994). *Turtles, termites, and traffic jams : explorations in massively parallel microworlds.* Cambridge, MA: MIT Press.

Reynolds, C. W. (1987). Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics, 21*(4), 25–34.

Rieskamp, J. (2001). *Simple strategies for social interactions: How to generate trust and fairness.* Unpublished doctoral dissertation, Free University of Berlin, Retrieved on 28 March, 2007, from http://www.diss.fu-berlin.de/cgi-bin/tar.cgi/2003/276/Fubdiss2003276.tar.gz.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of personality, 35*, 651–665.

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*(1), 1–7.

Sarkar, P. (2000). A brief history of cellular automata. *ACM Computing Surveys, 32*(1), 80–107.

Searle, J. R. (2001). *Rationality in action.* Cambridge, MA: MIT Press.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: the adaptive toolbox*. Cambridge, MA: MIT Press.

Shoemaker, P. J., Tankard, J. W. J., & Lasorsa, D. L. (2004). *How to build social science theories*. Thousand Oaks, CA: Sage.

Simon, H. A. (1978). Rationality as process and as product of thought. *American economic review*, *68*(2), 1–16.

Sloman, A. (1998). *What sort of architecture is required for a human-like agent*.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahnemann (Eds.), *Heuristics and biases: The psychology of intuitive judgement*. Cambridge: Cambridge University Press.

Small, C. (1999). Finding an invisible history: A computer simulation experiment (in Virtual Polynesia). *Journal of Artificial Societies and Social Simulation*, *2*(3), Retrieved 11 August, 2007, from http://www.soc.surrey.ac.uk/JASSS/2/3/6.html.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahnemann (Eds.), *Heuristics and biases: The psychology of intuitive judgement*. Cambridge: Cambridge University Press.

Stich, S., & Ravenscroft, I. (1994). What is folk psychology? *Cognition*, *50*, 447–468.

Strader, T. J., & Ramaswami, S. N. (2002). The value of seller trustworthiness in C2C online markets. *Communications of the ACM*, *45*(12), 45–49.

Suarez, M. (2003). Scientific representation: against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*(3), 225–244.

Sun, R. (2001). Cognitive science meets multi-agent systems: a prolegomenon. *Philosophical Psychology*, *14*(1), 5–24.

Suppe, F. (1989). *The semantic conception of theories and scientific realism*. Urbana: University of Illinois Press.

Sutton, R. M., & McClure, J. (2001). Covariational influences on goal-based explanation: an integrative model. *Journal of personality and social psychology*, *80*(2), 222–236.

Tobias, R., & Hofmann, C. (2004). Evaluation of free Java-libraries for social-scientific agent-based simulation. *Journal of Artificial Societies and Social Simulation*, *7*(1), Retrieved on 28 May, 2006, from http://jasss.soc.surrey.ac.uk/7/1/6.html.

Trade Me. (2007). *Site help: Learning more about the seller.* Retrieved 14 March 2007 from http://www.trademe.co.nz/Help.

Udehn, L. (2002). The changing face of methodological individualism. *Annual review of Sociology*, *28*, 479–507.

Ullmann-Margalit, E. (2002). Trust out of distrust. *Journal of Philosophy*, *99*(10), 532–548.

Vallacher, R. R., & Nowak, A. (1997). The emergence of dynamical social psychology. *Psychological Inquiry*, *8*(2), 73–99.

Van Hezewijk, R. (2004). Waking up from the dream of reason: Rationality

in the real world. *Theory and psychology*, *14*(2), 261–272.
Wilensky, U. (1999). *NetLogo.* Retrieved May 30, 2006, from http://ccl.northwestern.edu/netlogo/.
Williamson, O. E. (1993). Calculativeness, trust and economic organization. *Journal of law and economics*, *36*(1, Pt.2), 453–486.
Winsberg, E. (2001). Simulations, models and theories: Complex systems and their representations. *Philosophy of science*, *68*, S442–S454.
Wolfram, S. (2002). *A new kind of science*. Champaign, Il: Wolfram Media.
Wooldridge, M. (2002). *An introduction to multiagent systems.* Chichester: John Wiley.
Yamagishi, T. (2001). Trust as a form of social intelligence. In K. S. Cook (Ed.), *Trust in society.* New York: Russell Sage Foundation.
Yamagishi, T. (2003). Cross-societal experimentation on trust: a comparison of the United States and Japan. In E. Ostrom & J. Walker (Eds.), *Trust and reciprocity: Interdisciplinary lessons from experimental research* (pp. 352–370). New York: Russell Sage Foundation.
Yamagishi, T., Cook, K. S., & Watanabe, M. (1998). Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*, *104*(1), 165–194.
Zak, P. J. (2003). Trust. *Capco Institute Journal of Financial Transformation*, Retrieved on 17 July, 2003, from http://www.capco.com/journal.html.
Zak, P. J., & Knack, S. (2001). Trust and growth. *The Economic Journal*, *111*, 295–321.

Appendix A

Programme listings

Basic Breaking Model

BasicBreaking.java

```
package basicBreaking;
/**
* Basic Breaking Model
*
* Version 1
* @author Sue Street
* @version 1.0
*
* Created 23/2/07
*/
import java.util.ArrayList;
import java.text.NumberFormat;
import java.util.Iterator;
import uchicago.src.sim.analysis.DataRecorder;
import uchicago.src.sim.engine.BasicAction;
import uchicago.src.sim.engine.Schedule;
import uchicago.src.sim.engine.SimInit;
import uchicago.src.sim.engine.SimModelImpl;
import uchicago.src.sim.util.SimUtilities;
/**
* @author S.E.Street
```

```
* @version Minimal R&G 2003 with added relationship breakdown
* A bare cognitive model of trust from Rieskamp.
* Recording added recording to a file.
* Randomised added individual variation.
* Breaking adds a relationship split if either:
* the Investor fails to invest enough, or
* the Borrower fails to return enough.
*
* This code uses the RePast tutorial by J.T. Murphy as a foundation
* http://www.u.arizona.edu/~jtmurphy/H2R/main.htm
*/
public class BasicBreaking extends SimModelImpl{
/** Default values for the Rieskamp-Gigerenzer Model
*/
public static final double ITERATIONS = 10.0;
private double iterations = ITERATIONS;
private static final int HESITANT=1;
private static final int MODERATELY_GRIM=2;
private static final int REACTIVE=1;
private static final int HALF_BACK=2;
private int hesitant=HESITANT;
private int moderatelyGrim=MODERATELY_GRIM;
private int reactive=REACTIVE;
private int halfBack=HALF_BACK;
private static final int NUMAGENTS=100;
private int numAgents=NUMAGENTS;
private int numInvestor;
private int numBorrower;
private int numHesitant;
private int numReactive;
private static final double RECIPROCITY_MEAN = 0.34;
private double reciprocityThreshhold = RECIPROCITY_MEAN;
private static final double TRUST_THRESHHOLD_MEAN_REACTIVE =
0.17;
private double trustThreshholdReactive =
```

```
TRUST_THRESHHOLD_MEAN_REACTIVE;
private static final double TRUST_THRESHHOLD_MEAN_HALFBACK =
0.12;
private double trustThreshholdHalfBack =
TRUST_THRESHHOLD_MEAN_HALFBACK;
private double threshhold;
private int idInvestor;
private int strategyInvestor;
private double moneyInvestor;
private int stateInvestor;
private int idBorrower;
private int strategyBorrower;
private double moneyBorrower;
private int stateBorrower;
private Schedule schedule;
private ArrayList agentListInvestor;
private ArrayList agentListBorrower;
private DataRecorder recorder;
public String getName(){
return "Simple trust heuristic with breaking";
}
public void setup() {
agentListInvestor=new ArrayList();
agentListBorrower=new ArrayList();
schedule=new Schedule(1);
}
public void begin() {
buildModel();
buildSchedule();
}
private void buildModel() {
populate();
recorder=new DataRecorder("/home/sue/models/basicBreaking6/data.txt
", this);
recorder.createNumericDataSource("idInvestor", this,
```

```
"getIdInvestor");
recorder.createNumericDataSource("strategyInvestor",
this, "getStrategyInvestor");
recorder.createNumericDataSource("moneyInvestor",
this, "getMoneyInvestor");
recorder.createNumericDataSource("idBorrower", this,
"getIdBorrower");
recorder.createNumericDataSource("strategyBorrower", this, "getStrategy-
Borrower");
recorder.createNumericDataSource("moneyBorrower", this, "getMoneyBor-
rower");
}
private void buildSchedule() {
class TrustStep extends BasicAction{
public void execute(){
SimUtilities.shuffle(agentListInvestor);
Iterator e = agentListInvestor.iterator();
while (e.hasNext()){
TrustAgent tra=(TrustAgent)e.next();
int ID=tra.getID();
int p=tra.getPartner();
/**
* Check first whether the agent has a partner. If not, then try to find one.
*/
if (p==0){
p=findPartner();
TrustAgent trp = (TrustAgent)agentListBorrower.get(p);
tra.setPartner(p);
trp.setPartner(ID);
}
/**
* Only go through the transaction step with an agent that has a partner.
*/
if(p!=0){
TrustAgent trp = (TrustAgent)agentListBorrower.get(p-1);
```

```
int stateA = tra.getState();
int strategyA = tra.getStrategy();
double investMoney = tra.invest(strategyA, stateA);
int stateP = trp.getState();
int strategyP = trp.getStrategy();
double threshhold = 0;
switch(strategyP){
case 1:
threshhold = trp.getTrustThreshholdReactive();
break;
case 2:
threshhold = trp.getTrustThreshholdHalfBack();
break;
}
double back = trp.respond(investMoney, stateP, strategyP, threshhold);
int statePnew = trp.stateUpdate(investMoney, stateP, strategyP,
threshhold);
trp.setState(statePnew);
int stateAnew = tra.assess(back, investMoney, reciprocityThreshhold,
strategyA, stateA);
tra.setState(stateAnew);
boolean breakFlag = tra.assessRel(stateA, strategyA);
if (breakFlag){
tra.setPartner(0);
trp.setPartner(0);
tra.setState(0);
trp.setState(0);
}
idInvestor=tra.getID();
strategyInvestor=tra.getStrategy();
moneyInvestor=tra.getMoney();
idBorrower=trp.getID();
strategyBorrower=trp.getStrategy();
moneyBorrower=trp.getMoney();
recorder.record();
```

```
}
}
}
}
class RunEnd extends BasicAction {
public void execute(){
stop();
}
}
schedule.scheduleActionBeginning(0, new TrustStep());
schedule.scheduleActionBeginning(iterations,new RunEnd());
schedule.scheduleActionAtEnd(recorder,"writeToFile");
}
private void populate (){
/**Work out the number of investors and borrowers,
* then the number of Hesitant Investors and the number of Reactive
* Borrowers. The remainder of Investors are Moderately Grim and the
* remainder of Borrowers are Half Back.
*/
numInvestor=numAgents/2;
numHesitant=numInvestor/2;
int count=1;
threshhold = reciprocityThreshhold;
for(count=1; count<=numHesitant; count++){
addNewInvestor(hesitant,count,threshhold);
}
int c=count;
for(count=c; count<=numInvestor; count++){
addNewInvestor(moderatelyGrim,count,threshhold);
}
/*
* Adds the borrowers. The trust threshhold is different for the
* different borrower strategies.
*/
numBorrower=numAgents-numInvestor;
```

```
numReactive=numBorrower/2;
count=1;
threshhold = trustThreshholdReactive;
for(count=1; count<=numReactive; count++){
addNewBorrower(reactive,count,threshhold);
}
c=count;
threshhold = trustThreshholdHalfBack;
for(count=c; count<=numBorrower; count++){
addNewBorrower(halfBack,count,threshhold);
}
/**Should be fully populated by here.
*
* Now find partners. Shuffle the list of Investors, then work through
* the list pairing up the Investor with the corresponding Borrower
*
*/
SimUtilities.shuffle(agentListInvestor);
Iterator e = agentListInvestor.iterator();
while (e.hasNext()){
TrustAgent tra = (TrustAgent)e.next();
int ID=tra.getID();
int p=findPartner();
TrustAgent trp = (TrustAgent)agentListBorrower.get(p);
tra.setPartner(p);
trp.setPartner(ID);
}
}
private void addNewInvestor(int s, int id, double t){
int strategy=s;
int ID=id;
threshhold = t;
TrustAgent a = new TrustAgent(true, strategy, ID, threshhold);
agentListInvestor.add(a);
}
```

```
private void addNewBorrower(int s, int id, double t){
int strategy=s;
int ID=id;
threshhold = t;
TrustAgent a = new TrustAgent(false, strategy, ID, threshhold);
agentListBorrower.add(a);
}
public Schedule getSchedule() {
return schedule;
}
public String[] getInitParam(){
String[] initParams = {"Iterations", "NumAgents" , "ReciprocityThreshhold"
, "TrustThreshholdReactive" , "TrustThreshholdHalfBack"};
return initParams;
}
public double getIterations(){
return iterations;
}
public void setIterations(double it){
iterations=it;
}
public int getNumAgents(){
return numAgents;
}
public void setNumAgents(int na){
numAgents=na;
}
public double getReciprocityThreshhold(){
return reciprocityThreshhold;
}
public void setReciprocityThreshhold(double rt){
reciprocityThreshhold=rt;
}
public double getTrustThreshholdReactive(){
return trustThreshholdReactive;
```

```
}
public void setTrustThreshholdReactive(double ttr){
trustThreshholdReactive=ttr;
}
public double getTrustThreshholdHalfBack(){
return trustThreshholdHalfBack;
}
public void setTrustThreshholdHalfBack(double tth){
trustThreshholdHalfBack=tth;
}
public int getIdInvestor(){
return idInvestor;
}
public int getStrategyInvestor(){
return strategyInvestor;
}
public int getStateInvestor(){
return stateInvestor;
}
public double getMoneyInvestor(){
return moneyInvestor;
}
public int getIdBorrower(){
return idBorrower;
}
public int getStrategyBorrower(){
return strategyBorrower;
}
public int getStateBorrower(){
return stateBorrower;
}
public double getMoneyBorrower(){
return moneyBorrower;
}
public int findPartner(){
```

```
int newPartnerID=0;
Iterator e = agentListBorrower.iterator();
do{
TrustAgent trialP = (TrustAgent)e.next();
int checkPartner = trialP.getPartner();
switch (checkPartner){
case 0:
newPartnerID = trialP.getID()-1;
break;
default:
}
}
while((newPartnerID==0)&&(e.hasNext()));
return newPartnerID;
}
public static String format(double number, int frac){
NumberFormat N=NumberFormat.getInstance();
N.setGroupingUsed(false);
N.setMaximumFractionDigits(frac);
N.setMinimumFractionDigits(frac);
String num = N.format(number);
return num;
}
public static void main(String[] args) {
SimInit init = new SimInit();
BasicBreaking model = new BasicBreaking();
init.loadModel(model,"",false);
}
}
```

TrustAgent

```
package basicBreaking;
/**
* Basic Breaking Model
```

```
*
* Version 1
* @author Sue Street
* @version 1.0
*
* Created 23/2/07
*/
import uchicago.src.sim.util.Random;
import java.text.NumberFormat;
/**
* @author S.E.Street
* @version Minimal R&G 2003 with added relationship breakdown
*
* A bare cognitive model of trust from Rieskamp & Gigerenzer (2003).
* Recording added recording to a file.
* Randomised *removed*
* Breaking adds a relationship split if either:
* the Investor fails to invest enough, or
* the Borrower fails to return enough.
*
* This code uses the RePast tutorial by J.T. Murphy as a foundation
* http://www.u.arizona.edu/~jtmurphy/H2R/main.htm
*/
public class TrustAgent {
/**Transaction variables declared here
*
*/
private double dividend;
private static final double INCOME = 10;
private double income = INCOME;
private double invest;
private double investMoney;
private double money;
//Original R&G figure of 3.0
private static final double PROFIT_MARGIN = 3.0;
```

```
private double profitMargin = PROFIT_MARGIN;
private double proportionInvested;
/** Personality variables declared here
*
* Variables setting bottom line for fairness and trust.
* Default for investor is at least the investment returned for reciprocity.
* Default for borrowers is at least 12% (half-back) or 17% (reactive) for
trust
* to be acknowledged.
*
* Now has provision for mean and standard deviation to be fixed for the
population
* with individual values drawn from a distribution.
*/
private static final double RECIPROCITY_SD = 0.0;
private static final double TRUST_THRESHHOLD_SD_REACTIVE = 0.0;
private static final double TRUST_THRESHHOLD_SD_HALFBACK = 0.0;
private double threshhold;
private double reciprocityThreshhold;
private double trustThreshholdReactive;
private double trustThreshholdHalfBack;
/**
* Variables setting investment levels and responses for each state
*/
public static final double INVEST_HESITANT_MEAN_1 = 0.5;
public static final double INVEST_HESITANT_SD_1 = 0.0;
public double investHesitant1 = INVEST_HESITANT_MEAN_1;
public static final double INVEST_HESITANT_MEAN_2 = 1.0;
public static final double INVEST_HESITANT_SD_2 = 0.0;
public double investHesitant2 = INVEST_HESITANT_MEAN_2;
public static final double INVEST_HESITANT_MEAN_3 = 0.00;
public static final double INVEST_HESITANT_SD_3 = 0.00;
public double investHesitant3 = INVEST_HESITANT_MEAN_3;
private static final double INVEST_MOD_GRIM_MEAN_1 = 1.0;
private static final double INVEST_MOD_GRIM_SD_1 = 0.0;
```

```
private double investModGrim1 = INVEST_MOD_GRIM_MEAN_1;
private static final double INVEST_MOD_GRIM_MEAN_2 = 1.0;
private static final double INVEST_MOD_GRIM_SD_2 = 0.0;
private double investModGrim2 = INVEST_MOD_GRIM_MEAN_2;
private static final double INVEST_MOD_GRIM_MEAN_3 = 0.0;
private static final double INVEST_MOD_GRIM_SD_3 = 0.0;
private double investModGrim3 = INVEST_MOD_GRIM_MEAN_3;
private static final double RETURN_REACT_MEAN_1 = 0.0;
private static final double RETURN_REACT_SD_1 = 0.0;
private double returnReact1 = RETURN_REACT_MEAN_1;
private static final double RETURN_REACT_MEAN_2 = 0.7;
private double returnReact2 = RETURN_REACT_MEAN_2;
private static final double RETURN_REACT_SD_2 = 0.0;
private static final double RETURN_HALF_BACK_MEAN_1 = 0.5;
private static final double RETURN_HALF_BACK_SD_1 = 0.0;
private double returnHalfBack1 = RETURN_HALF_BACK_MEAN_1;
private static final double RETURN_HALF_BACK_MEAN_2 = 0.0;
private static final double RETURN_HALF_BACK_SD_2 = 0.0;
private double returnHalfBack2 = RETURN_HALF_BACK_MEAN_2;
/** Strategy variables
*
*/
private static final boolean BORROWER = false;
private boolean isInvestor = BORROWER;
private int strategy;
/** Identity variables
*
*/
private int ID;
private int partner;
/**Transaction state variables
*
*/
private static final int STATE = 1;
private int state = STATE;
```

```
/**
* Initialise a Trust Agent. It begins with no money and no partner. It draws
a
* role and strategy from the populate method.
* @param strategy
* @param isInvestor
*/
public TrustAgent(boolean iI, int s, int id, double t) {
ID = id;
money = 0;
partner = 0;
isInvestor = iI;
strategy = s;
threshhold = t;
/**
* Fire up a random number generator for a Normal distribution mean = 0,
sd = 1
*/
Random.createNormal(0,1);
/**
* Then randomly allocate individual characteristics, based on a mean and
SD
* as specified in the constants at the beginning.
*/
if (isInvestor==true){
double rand=Random.normal.nextDouble();
reciprocityThreshhold=threshhold+(rand*RECIPROCITY_SD);
if (strategy==1){
rand=Random.normal.nextDouble();
investHesitant1=INVEST_HESITANT_MEAN_1+
(rand*INVEST_HESITANT_SD_1);
rand=Random.normal.nextDouble();
investHesitant2=INVEST_HESITANT_MEAN_2
+(rand*INVEST_HESITANT_SD_2);
rand=Random.normal.nextDouble();
investHesitant3=INVEST_HESITANT_MEAN_3
```

```
+(rand*INVEST_HESITANT_SD_3);
}
else{
rand=Random.normal.nextDouble();
investModGrim1=INVEST_MOD_GRIM_MEAN_1
+(rand*INVEST_MOD_GRIM_SD_1);
rand=Random.normal.nextDouble();
investModGrim2=INVEST_MOD_GRIM_MEAN_2
+(rand*INVEST_MOD_GRIM_SD_2);
rand=Random.normal.nextDouble();
investModGrim3=INVEST_MOD_GRIM_MEAN_3
+(rand*INVEST_MOD_GRIM_SD_3);
}
}
else{
// These are the borrowers
double rand;
if(strategy==1){
rand=Random.normal.nextDouble();
trustThreshholdReactive=threshhold
+rand*TRUST_THRESHHOLD_SD_REACTIVE;
rand=Random.normal.nextDouble();
returnReact1=RETURN_REACT_MEAN_1
+(rand*RETURN_REACT_SD_1);
rand=Random.normal.nextDouble();
returnReact2=RETURN_REACT_MEAN_2
+(rand*RETURN_REACT_SD_2);
}
else{
rand=Random.normal.nextDouble();
trustThreshholdHalfBack=threshhold
+rand*TRUST_THRESHHOLD_SD_HALFBACK;
rand=Random.normal.nextDouble();
returnHalfBack1=RETURN_HALF_BACK_MEAN_1
+(rand*RETURN_HALF_BACK_SD_1);
```

```
rand=Random.normal.nextDouble();
returnHalfBack2=RETURN_HALF_BACK_MEAN_2
+(rand*RETURN_HALF_BACK_SD_2);
}
}
}
/**
* Allow model to access money held by this agent
* @return money
*/
public double getMoney() {
return money;
}
/**
* Allow model to access this agent's strategy
* @return strategy
*/
public int getStrategy() {
return strategy;
}
/**
* Allow model to access role (Investor or Borrower)
* @return isInvestor
*/
public boolean getRole() {
return isInvestor;
}
/**
* Allow model to identify this agent
* @return ID
*/
public int getID() {
return ID;
}
/**
```

```
* Allow model to identify this agent's partner
* @return partner
*/
public int getPartner() {
return partner;
}
/**
* Allow model to set the agent's partner in initialisation
* @return partner
*/
public void setPartner(int p) {
partner = p;
}
/**
* Allow model to access and set the agent's state
*/
public int getState(){
return state;
}
public void setState(int s){
state=s;
}
public double getTrustThreshholdReactive(){
return trustThreshholdReactive;
}
public double getTrustThreshholdHalfBack(){
return trustThreshholdHalfBack;
}
public double getReciprocityThreshhold(){
return reciprocityThreshhold;
}
/**
* Invest carries out the first step of the interaction cycle. It activates a method
* for either a Hesitant (investHesitant) or a Moderately grim (investModeratelyGrim)
```

```
* investor.
*
* @return Amount to be invested with the borrower
*/
public double invest(int strat, int s) {
strategy = strat;
state = s;
switch (strategy) {
case 1 :
investMoney = investHesitant(state);
break;
case 2 :
investMoney = investModeratelyGrim(state);
}
return investMoney;
}
/**
* The Hesitant Investor decides how much to invest
* @return investedMoney
*/
public double investHesitant(int s) {
state = s;
switch (state) {
case 1 :
invest = investHesitant1;
break;
case 2 :
invest = investHesitant2;
break;
case 3 :
invest = investHesitant3;
}
double invMon = invest * income;
return invMon;
}
```

```
/**
* The Moderately Grim Investor decides how much to invest
* @return investedMoney
*/
private double investModeratelyGrim(int s) {
int state = s;
switch (state) {
case 1 :
invest = investModGrim1;
break;
case 2 :
invest = investModGrim2;
break;
case 3 :
invest = investModGrim3;
}
double invMon = invest * income;
return invMon;
}
/**
* Once the Investor has completed the investment decision, the Borrower
* assesses whether or not it has been trusted, and on the basis of this and
* its current state the Borrower decides how much to return to the Investor.
* The difference between the amount received and the dividend returned is
added
* to the Borrower's wealth
* @param investedMoney
* @return dividend
*/
public double respond(double inv, int s, int strat, double t) {
investMoney = inv;
state = s;
strategy = strat;
threshhold = t;
double received = profitMargin * investMoney;
```

```
proportionInvested = investMoney / income;
switch (strategy) {
case 1 :
dividend = respondReactive(received, proportionInvested, threshhold);
break;
case 2 :
dividend = respondHalfBack(received, proportionInvested, threshhold,
state);
}
double kept = received - dividend;
money += kept;
return dividend;
}
public int stateUpdate(double inv, int s, int strat, double t) {
investMoney = inv;
state = s;
strategy = strat;
double threshholdH = t;
proportionInvested = investMoney / income;
switch (strategy) {
case 1 :
state = 1;
break;
case 2 :
state = stateHalfBack(proportionInvested, threshholdH, state);
}
return state;
}
/**
* @param received
* @return dividend
*/
private double respondReactive(double r, double pI, double tTR) {
double received = r;
double proportionInvested = pI;
```

```
double trustThreshholdReactive = tTR;
double d = 0;
if (proportionInvested >= trustThreshholdReactive) {
d = returnReact2;
} else {
d = returnReact1;
}
dividend = d * received;
return dividend;
}
/**
* @param received
* @return dividend
*/
private double respondHalfBack(double r, double pI, double tTH, int s) {
double received = r;
double proportionInvested = pI;
double threshholdH = tTH;
int state = s;
boolean t = true;
if (proportionInvested < threshholdH) {
t = false;
}
if (state == 1) {
if (t) {
dividend = received * returnHalfBack1;
} else {
dividend = received * returnHalfBack2;
}
}
if (state == 2) {
dividend = received * returnHalfBack1;
}
return dividend;
}
```

```
private int stateHalfBack(double pI, double tTH, int s) {
double proportionInvested = pI;
double threshholdH = tTH;
int state = s;
boolean t = true;
if (proportionInvested < threshholdH) {
t = false;
}
if (state == 1) {
if (t) {
state = 1;
} else {
state = 2;
}
}
if (state == 2) {
state = 1;
}
return state;
}
/**
*
*/
public int assess(double b, double i, double rT, int strat, int s) {
double back = b;
double investMoney = i;
double reciprocityThreshhold = rT;
int state = s;
int strategy = strat;
double returnRatio = back / investMoney;
boolean reciprocated = false;
if (returnRatio >= reciprocityThreshhold) {
reciprocated = true;
}
switch (strategy) {
```

```
case 1 :
state = assessHesitant(reciprocated, state);
break;
case 2 :
state = assessModeratelyGrim(reciprocated, state);
}
money += back;
return state;
}
/**
* Investor assesses whether the investment was reciprocated and adjusts state
* if a change is required
*
* @param reciprocated
* @param state
* @return state
*/
private int assessModeratelyGrim(boolean r, int s) {
int state = s;
boolean reciprocated = r;
if (state == 1) {
if (reciprocated) {
state = 2;
} else {
state = 3;
}
}
if (state == 2) {
if (reciprocated)
state = 2;
else
state = 1;
}
if (state == 3)
```

```
state = 3;
return state;
}
private int assessHesitant(boolean r, int s) {
state = s;
boolean reciprocated = r;
if (state == 1)
state = 2;
if (state == 2) {
if (reciprocated)
state = 2;
else
state = 3;
}
if (state == 3)
state = 1;
return state;
}
public boolean assessRel(int s, int strat){
int state=s;
int strategy = strat;
boolean breakFlag = false;
if (state == 3){
switch(strategy){
case 1:
breakFlag = false;
break;
case 2:
breakFlag = true;
}
}
return breakFlag;
}
public static String form(double number, int frac){
NumberFormat N = NumberFormat.getInstance();
```

```
N.setMaximumFractionDigits(frac);
N.setMinimumFractionDigits(frac);
String num = N.format(number);
return num;
}
}
```

Trading model

TradeMe11.java

```
package tradeMe11;
/**
* TradeMe Version 11. File output to
* file output<tick#>.txt at each tick.
* Fixes V10 problem with the information exchange
* in the strategy sharing - was not sharing it into
* the correct variable.
*
* Version 11
* @author Sue Street
* @version 11.0
*
*/
import java.io.IOException;
import java.util.ArrayList;
import java.util.Iterator;
import uchicago.src.sim.engine.BasicAction;
import uchicago.src.sim.engine.Schedule;
import uchicago.src.sim.engine.SimInit;
import uchicago.src.sim.engine.SimModelImpl;
import uchicago.src.sim.util.SimUtilities;
public class TradeMe11 extends SimModelImpl{
private double buyPrice = 0.0;
private double happyGoodTrade = 10.0;
```

```
private double happySuccessfulCon = 20.0;
private double unhappyBadComms = 2.0;
private double unhappyBadDelivery = 20.0;
private double unhappyBadPayment = 5.0;
private double neutralHappy = 0.0;
private double sellerHappiness = 0.0;
private double buyerHappiness = 0.0;
private int tradersNumAgents = 100;
private int ID;
private Schedule schedule;
private ArrayList agentListTraders;
/**
* Sets name as required by SimpleModel
*/
public String getName(){
return "Trade Me Version 11";
}
/**
* Initialises model, tearing down the agents and schedules from
* previous runs. Required by SimpleModel
*/
public void setup(){
agentListTraders = new ArrayList();
schedule = new Schedule(1);
}
/**
* Builds the model and schedule for this run.
* @throws IOException
*/
public void begin(){
buildModel();
buildSchedule();
}
/**
```

```
* Adds agents to the model. Initialises and sets variables for the Data Recorder
*/
private void buildModel(){
populate();
}
/**
* Schedule a sale round at each iteration
* and data output at the end of every 100th run.
* Stop the simulation after 500 runs.
*/
private void buildSchedule(){
class TradeStep extends BasicAction{
public void execute(){
sale();
try {
outputResult();
} catch (IOException e) {
}
}
}
class OutputResult extends BasicAction {
public void execute(){
try {
outputResult();
} catch (IOException e) {
}
}
}
class EndOfRun extends BasicAction{
public void execute(){
stop();
}
}
schedule.scheduleActionBeginning(0, new TradeStep());
```

```
schedule.scheduleActionAtInterval(100, new OutputResult());
schedule.scheduleActionAt(500, new EndOfRun());
}
/**
* Shuffle the agents then step through each agent to see if
* they have something to sell. Then for each seller step
* through each buyer for bids. Buyers cannot see previous
* bid. Seller may reject buyers, based on reputation.
*/
private void sale(){
SimUtilities.shuffle(agentListTraders);
Iterator iterateSeller = agentListTraders.iterator();
while (iterateSeller.hasNext()) {
Trader seller = (Trader) iterateSeller.next();
/*
* Each agent is polled to see if they have something to sell. If the asking
price
* is greater than zero, then the selling price is initialised and the bid pro-
cess
* starts. If the agent has nothing to sell then jump to the next agent.
*/
double askingPrice = seller.forSalePrice();
Trader buyer = seller;
/*
* Collect the seller's trading history to pass to the buyer.
*/
int bD = seller.getBadDelivery();
int bC = seller.getBadCommunication();
int bP = seller.getBadPayment();
int gT = seller.getGoodTrade();
int tT = seller.getTotalTrades();
boolean haveBuyer = false;
if (askingPrice > 0.0) {
double sellingPrice = 0.0;
/*
```

```
* If there is an item for sale then step through all agents to obtain bids.
* If the bid beats the previous highest bid, this becomes the highest bid.
* Currently does not treat asking price as a reserve price.
*/
SimUtilities.shuffle(agentListTraders);
Iterator iterateBuyer = agentListTraders.iterator();
while (iterateBuyer.hasNext()) {
Trader bidder = (Trader) iterateBuyer.next();
/*
* Buyer decides how much to bid and whether to bid based on the
* trading history
*/
double bid = bidder.bidToBuy(bD, bC, bP, gT, tT);
if (bid > sellingPrice) {
sellingPrice = bid;
buyer = bidder;
buyer.setBoughtInLastRound(true);
haveBuyer = true;
seller.setTotalTrades(seller.getTotalTrades()+1);
buyer.setTotalTrades(buyer.getTotalTrades()+1);
// double lap = getTickCount();
// System.out.println(lap+" "+seller.getId()+" "+seller.getTotalTrades()+" "
+buyer.getId()+" "+buyer.getTotalTrades());
}
}
if (haveBuyer) {
saleResult(buyer, seller);
buyer.exchangeStrategy(seller);
seller.exchangeStrategy(buyer);
}
}
}
}
/**
```

```
* Assess outcome of the sale for buyer and seller. Seller is happy if trade is completed
* or if buyer pays and seller does not deliver. If either fails to communicate, seller fails
* to deliver or buyer fails to pay then other party is unhappy according to parameters set,
* and the failure to perform is recorded against the agent that fails to meet obligations.
*
* @param buyer
* @param seller
*/
private void saleResult(Trader buyer, Trader seller){
if (!buyer.communicate()) {
// Seller is unhappy at the lack of communication
sellerHappiness = -unhappyBadComms;
seller.setOverallHappiness(seller.getOverallHappiness()+sellerHappiness);
// Buyer is neither happy nor unhappy, and gets a bad communication recorded
buyerHappiness = neutralHappy;
buyer.setHappyWithTransaction(buyerHappiness);
buyer.setOverallHappiness(buyer.getOverallHappiness() + buyerHappiness);
buyer.setBadCommunication(buyer.getBadCommunication() + 1);
} else if (!seller.communicate()) {
// Seller is neither happy nor unhappy, and gets a bad communication recorded
sellerHappiness = neutralHappy;
seller.setOverallHappiness(seller.getOverallHappiness()+sellerHappiness);
int noBadComms = seller.getBadCommunication() + 1;
seller.setBadCommunication(noBadComms);
// Buyer is unhappy at the lack of communication
buyerHappiness = -unhappyBadComms;
buyer.setHappyWithTransaction(buyerHappiness);
buyer.setOverallHappiness
(buyer.getOverallHappiness() + buyerHappiness);
} else if (!buyer.payUp()) {
```

```
// Seller is unhappy at lack of payment
sellerHappiness = -unhappyBadPayment;
seller.setOverallHappiness(seller.getOverallHappiness()+sellerHappiness);
//Buyer is neither happy nor unhappy, but gets a bad payment report
buyerHappiness = neutralHappy;
buyer.setOverallHappiness(buyer.getOverallHappiness() + buyerHappiness);
buyer.setBadPayment(buyer.getBadPayment() + 1);
} else if (!seller.delivery()) {
// Seller is happy as payment received but no goods delivered
sellerHappiness = happySuccessfulCon;
seller.setOverallHappiness(seller.getOverallHappiness()+sellerHappiness);
seller.setBadDelivery(seller.getBadDelivery() + 1);
// Buyer is unhappy as payment was paid but no goods received
buyerHappiness = -unhappyBadDelivery;
buyer.setHappyWithTransaction(buyerHappiness);
buyer.setOverallHappiness(buyer.getOverallHappiness() + buyerHappiness);
} else {
// The trade goes through cleanly - both buyer and seller are happy with a
good trade
sellerHappiness = happyGoodTrade;
seller.setOverallHappiness(seller.getOverallHappiness()+sellerHappiness);
buyerHappiness = happyGoodTrade;
buyer.setHappyWithTransaction(buyerHappiness);
buyer.setOverallHappiness(buyer.getOverallHappiness() + buyerHappiness);
buyer.setGoodTrade(buyer.getGoodTrade() + 1);
seller.setGoodTrade(seller.getGoodTrade() + 1);
}
}
/**
* Fill the simulation with the number of agents set in the
* parameters.
*/
private void populate(){
for(int i=1; i<=tradersNumAgents; i++){
addNewTrader(i);
```

```
}
}
private void addNewTrader(int id){
ID=id;
Trader a = new Trader(ID);
agentListTraders.add(a);
a.init();
}
public Schedule getSchedule(){
return schedule;
}
public void outputResult() throws IOException{
double time=getTickCount()+1000;
Stream fout = new Stream("output"+time+".txt");
fout.println(" Happiness, Reliability, Honesty, badComms,
badPay, badDel, goodTrade, "+
"TotTrades, WeightBadComms, WeightBadPay, WeightBadDel, WeightGood-
Trade, WeightTotTrades");
Iterator iterator = agentListTraders.iterator();
while (iterator.hasNext()){
Trader trader = (Trader) iterator.next();
int tID = trader.getId();
String stID = Stream.format(tID, 3);
double hWT = trader.getOverallHappiness();
String shWT = Stream.format(hWT, 9, 0);
double r = trader.getReliability();
String sr = Stream.format(r, 11, 4);
double hH = trader.getHowHonest();
String shH = Stream.format(hH, 7, 4);
int bC = trader.getBadCommunication();
String sbC = Stream.format(bC, 8);
int bP = trader.getBadPayment();
String sbP = Stream.format(bP, 6);
int bD = trader.getBadDelivery();
String sbD = Stream.format(bD, 6);
```

```
int gT = trader.getGoodTrade();
String sgT = Stream.format(gT, 9);
double gTT = trader.getTotalTrades();
String sgTT = Stream.format(gTT, 9, 1);
double wBC = trader.getWeightBadComm();
String swBC = Stream.format(wBC, 14, 4);
double wBP = trader.getWeightBadPay();
String swBP = Stream.format(wBP, 12, 4);
double wBD = trader.getWeightBadDel();
String swBD = Stream.format(wBD, 12, 4);
double wGT = trader.getWeightGoodTrade();
String swGT = Stream.format(wGT, 15, 6);
double wTT = trader.getWeightTotalTrades();
String swTT = Stream.format(wTT, 15, 4);
fout.println(stID+", "+shWT+", "+sr+", "+shH+", "+sbC+", "+sbP+", "+sbD+", "+sgT
+", "+sgTT+", "+swBC+", "+swBP+", "+swBD+", "+swGT+", "+swTT);
}
fout.close();
}
public String[] getInitParam(){
String[] initParams = {"tradersNumAgents","unhappyBadComms",
"unhappyBadDelivery", "unhappyBadPayment",
"neutralHappy" , "happyGoodTrade" , "happySuccessfulCon"
};
return initParams;
}
public void setBuyPrice(double d) {
buyPrice = d;
}
public double getBuyPrice() {
return buyPrice;
}
public void setTradersNumAgents(int i) {
tradersNumAgents = i;
```

```
}
public int getTradersNumAgents() {
return tradersNumAgents;
}
public double getUnhappyBadComms() {
return unhappyBadComms;
}
public void setUnhappyBadComms(double d) {
unhappyBadComms = d;
}
public double getUnhappyBadDelivery() {
return unhappyBadDelivery;
}
public void setUnhappyBadDelivery(double d) {
unhappyBadDelivery= d;
}
public double getUnhappyBadPayment() {
return unhappyBadPayment;
}
public void setUnhappyBadPayment(double d) {
unhappyBadPayment= d;
}
public double getNeutralHappy() {
return neutralHappy;
}
public void setNeutralHappy(double d) {
neutralHappy = d;
}
public double getHappyGoodTrade() {
return happyGoodTrade;
}
public void setHappyGoodTrade(double d) {
happyGoodTrade = d;
}
public double getHappySuccessfulCon() {
```

```
return happySuccessfulCon;
}
public void setHappySuccessfulCon(double d) {
happySuccessfulCon = d;
}
public static void main(String[] args) throws IOException {
SimInit init = new SimInit();
TradeMe11 model = new TradeMe11();
init.loadModel(model,"",false);
}
}
```

Trader.java

```
package tradeMe11;
/**
* TradeMe Version 11. File output to
* file output<tick#>.txt at each tick.
* Fixes V10 problem with the information exchange
* in the strategy sharing - was not sharing it into
* the correct variable.
*
* Version 11
* @author Sue Street
* @version 11.0
*
*/
import uchicago.src.sim.util.Random;
public class Trader
{
private double askingPrice;
private double biddingPrice;
private double reliability;
private boolean honesty;
private double tradeResult;
```

```
private int badDelivery;
private int badPayment;
private int badCommunication;
private double happyWithTransaction;
private double howHonest;
private int goodTrade;
private int totalTrades;
private boolean boughtInLastRound;
private double overallHappiness;
private int ID;
/* TradeMe8 adds an interaction so that traders can exchange
* weighting information after a successful trade. There is no
* exchange of information after an unsuccessful trade.
*
* This is probably less realistic than that traders exchange
* strategic information with their friends, but the essence remains
* that the information is adopted from more successful traders.
*/
private double weightBadComm = 0.0;
private double weightBadDel = 0.0;
private double weightBadPay = 0.0;
private double weightGoodTrade = 0.0;
private double weightTotalTrades = 0.0;
public Trader(int id) {
ID=id;
askingPrice = 0.0;
biddingPrice = 0.0;
reliability = 1.0;
honesty = true;
tradeResult = 0.0;
badDelivery = 0;
badCommunication = 0;
happyWithTransaction = 0.0;
howHonest = 0.9;
goodTrade = 0;
```

```
totalTrades = 0;
boughtInLastRound = false;
overallHappiness = 0.0;
/* New variables initialised to zero. They will be assigned a
* random weighting in the initialisation process.
*/
weightBadComm = 0.0;
weightBadDel = 0.0;
weightBadPay = 0.0;
weightGoodTrade = 0.0;
weightTotalTrades = 0.0;
}
public void setAskingPrice(double d) {
askingPrice = d;
}
public double getAskingPrice() {
return askingPrice;
}
public void setReliability(double d) {
reliability = d;
}
public double getReliability() {
return reliability;
}
public void setHonesty(boolean bool) {
honesty = bool;
}
public boolean getHonesty() {
return honesty;
}
public void setTradeResult(double d) {
tradeResult = d;
}
public double getTradeResult() {
return tradeResult;
```

```
}
public void setBadDelivery(int i) {
badDelivery = i;
}
public int getBadDelivery() {
return badDelivery;
}
public void setBadPayment(int i) {
badPayment = i;
}
public int getBadPayment() {
return badPayment;
}
public void setBadCommunication(int i) {
badCommunication = i;
}
public int getBadCommunication() {
return badCommunication;
}
public void setHappyWithTransaction(double d) {
happyWithTransaction = d;
}
public double getHappyWithTransaction() {
return happyWithTransaction;
}
public void setHowHonest(double d) {
howHonest = d;
}
public double getHowHonest() {
return howHonest;
}
public void setGoodTrade(int i) {
goodTrade = i;
}
public int getGoodTrade() {
```

```
return goodTrade;
}
public void setTotalTrades(int i){
totalTrades = i;
}
public int getTotalTrades(){
return totalTrades;
}
public void setBoughtInLastRound(boolean bool) {
boughtInLastRound = bool;
}
public boolean getBoughtInLastRound() {
return boughtInLastRound;
}
public void setOverallHappiness(double d) {
overallHappiness = d;
}
public double getOverallHappiness() {
return overallHappiness;
}
public void setId(int i) {
ID = i;
}
public int getId() {
return ID;
}
public double getWeightBadComm(){
return weightBadComm;
}
public void setWeightBadComm(double d){
weightBadComm=d;
}
public double getWeightBadDel(){
return weightBadDel;
}
```

```
public void setWeightBadDel(double d){
weightBadDel=d;
}
public double getWeightBadPay(){
return weightBadPay;
}
public void setWeightBadPay(double d){
weightBadPay=d;
}
public double getWeightGoodTrade(){
return weightGoodTrade;
}
public void setWeightGoodTrade(double d){
weightGoodTrade=d;
}
public double getWeightTotalTrades(){
return weightTotalTrades;
}
public void setWeightTotalTrades(double d){
weightTotalTrades=d;
}
public double forSalePrice() {
Random.createNormal((double) 20, (double) 5);
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
if (!(d > 0.5))
askingPrice = 0.0;
else
askingPrice = Random.normal.nextDouble();
return askingPrice;
}
public double bidToBuy(int bD, int bC, int bP, int gT, int tT) {
double badD = (double)bD / (double)tT;
double badC = (double)bC / (double)tT;
double badP = (double)bP / (double)tT;
```

```
double goodT= (double)gT / (double)tT;
double sellerAssessment = (weightTotalTrades*totalTrades
+weightGoodTrade*goodT
-weightBadComm*badC
-weightBadPay*badP
-weightBadDel*badD);
if(sellerAssessment<0.0){
biddingPrice=0.0;
}
else{
Random.createNormal((double) 20, (double) 5);
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
if (!(d > 0.5))
biddingPrice = 0.0;
else
biddingPrice = Random.normal.nextDouble();
}
return biddingPrice;
}
public void init() {
askingPrice = 0.0;
biddingPrice = 0.0;
Random.createUniform((double) 0, (double) 1);
reliability = Random.uniform.nextDouble();
/*
* Fourth root the reliability to make the distribution
* more weighted to high reliability
*/
reliability = Math.pow(reliability, 0.25);
howHonest = Random.uniform.nextDouble();
/*
* Fourth root the honesty to make the distribution more
* weighted to high honesty
*/
```

```
howHonest = Math.pow(howHonest, 0.25);
/*
* Assign a random weight of each reputation element for each agent.
* NOTE this is set up for 100 runs; the weighting on the number of
* of total trades being weighted 1/100th
*
* Each of these random elements adds a set of assumptions - assumptions
* in particular about the distribution of characteristics in a population.
* This tends to be something that is relatively unreported.
*/
weightBadComm = Random.uniform.nextDouble();
weightBadDel = Random.uniform.nextDouble();
weightBadPay = Random.uniform.nextDouble();
weightGoodTrade = Random.uniform.nextDouble();
weightTotalTrades = Random.uniform.nextDouble()/100.0;
}
public boolean payUp() {
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
boolean bool;
if (d >= reliability)
bool = false;
else
bool = true;
return bool;
}
public boolean communicate() {
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
boolean bool;
if (d >= reliability)
bool = false;
else
bool = true;
return bool;
```

```
}
public boolean delivery() {
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
boolean bool;
if (d >= howHonest)
bool = false;
else
bool = true;
return bool;
}
/**Randomly exchange information with partners reporting a
* greater level of happiness upon completion of a successful
* trade. The less happy trader has a 50% chance of adopting
* each individual element of the strategy.
*
* @param partner
*/
public void exchangeStrategy(Trader partner){
if (overallHappiness<partner.getOverallHappiness()){
double altWbC=partner.getWeightBadComm();
double altWbD=partner.getWeightBadDel();
double altWbP=partner.getWeightBadPay();
double altWgT=partner.getWeightGoodTrade();
double altWtT=partner.getWeightTotalTrades();
Random.createUniform((double) 0, (double) 1);
double d = Random.uniform.nextDouble();
if (d>0.5){
weightBadComm = altWbC;
}
d = Random.uniform.nextDouble();
if (d>0.5){
weightBadPay = altWbP;
}
d = Random.uniform.nextDouble();
```

```
if (d>0.5){
weightBadDel = altWbD;
}
d = Random.uniform.nextDouble();
if (d>0.5){
weightGoodTrade = altWgT;
}
d = Random.uniform.nextDouble();
if (d>0.5){
weightTotalTrades = altWtT;
}
}
}
public void setModel(TradeMe11 trademe) {
}
}
```

Stream.java

```
package tradeMe11;
import java.io.FileWriter;
import java.io.IOException;
import java.io.PrintWriter;
import java.text.DecimalFormat;
public class Stream {
/* Adapted from The Stream class by J M Bishop and B Worrall
* May 2000
*
* Constructors
* ________
* public Stream (String filename)
*
* Output
* ____
* public void println - for Objects, String, int, double, char
```

```
* public void print - for Objects, String, int, double, char
* public void close()
*
* Output - class methods
* ______________
* public String format (int number, int align)
* public String format (double number, int align, int frac)
*/
private PrintWriter out;
public Stream(String filename) throws IOException {
out = create(filename);
}
private PrintWriter create(String filename) throws IOException {
return new PrintWriter(new FileWriter(filename));
}
private static DecimalFormat N = new DecimalFormat();
private static final String spaces = " ";
public static String format(double number, int align, int frac) {
N.setGroupingUsed(false);
N.setMaximumFractionDigits(frac);
N.setMinimumFractionDigits(frac);
String num = N.format(number);
if (num.length() < align)
num = spaces.substring(0,align-num.length()) + num;
return num;
}
public static String format(int number, int align) {
N.setGroupingUsed(false);
N.setMaximumFractionDigits(0);
String num = N.format(number);
if (num.length() < align)
num = spaces.substring(0,align-num.length()) + num;
return num;
}
public void println(Object s) {
```

```
out.println(String.valueOf(s));
out.flush();
}
public void println(String s) {
out.println(s);
out.flush();
}
public void println(int s) {
out.println(s);
out.flush();
}
public void println(double s) {
out.println(s);
out.flush();
}
public void println(char s) {
out.println(s);
out.flush();
}
public void print(Object s) {
out.print(String.valueOf(s));
out.flush();
}
public void print(String s) {
out.print(s);
out.flush();
}
public void print(int s) {
out.print(s);
out.flush();
}
public void print(double s) {
out.print(s);
out.flush();
}
```

```
public void print(char s) {
out.print(s);
out.flush();
}
public void close() throws IOException {
if (out != null)
out.close();
}
public void flush() {
out.flush();
}
}
```