# Chapter 1

# Introduction

## 1.1 Overview

### 1.1.1 Agents and MAS

Much recent work in Distributed Artificial Intelligence (DAI) [97] has applied an "Agent Based" framework. The idea is that complex software systems can be interpreted, and implemented, as interacting systems of units termed "agents". Agents are software systems which have their own goals. They attempt to achieve their goals by selecting appropriate actions. Although there are many definitions of the term "agent", essentially all concede that agents have:

- A sensory mechanism (percepts or inputs) which can deliver information about the environment external to the agent.
- A decision mechanism which can select alternative forms of action.
- Effectors or output mechanisms that can effect the environment external to the agent.

Goals may be implicit in the decision mechanism or explicitly represented in the agent. Agents may have no internal state and simply react immediately to environmental inputs (reactive agents) or have explicitly stored beliefs which are updated based on sensory inputs and used to implement "human like" reasoning (cognitive agents).

Agents generally need to interact with other agents in order to achieve their goals. The key idea behind the "agent based paradigm" is that, rather than construct large complex single agents to solve particular problems, many simpler and smaller agents can be constructed which work together to solve complex problems by achieving their simpler individual goals. From this perspective there is much interest in protocols and mechanisms by which many agents, in the form of a Muti-Agent System (MAS), may co-operate to achieve their goals. To this end three major modes of investigation can be identified:

- Engineer actual working systems.
- Produce formal models of systems.
- Simulate and empirically investigate systems.

Because of the complexity of MAS the engineering route is still highly experimental. There are currently no mature engineering methodologies which address the MAS level of complexity. Consequently, practical working systems are rare. Those systems that are constructed are, for-the-most-part, experimental and ad-hoc. Formal models require specialised and novel logics and currently appear to have limited applicability, especially when agents are highly adaptive (that is, change their decision mechanisms based on experience). Formal models also have little to say about the emergence (see later) of macro structures (i.e. system level properties) from micro behaviours (i.e. individual agent actions) in MAS. Many researchers have therefore used simulation to explore complex agent interactions in an attempt to build agent theory. Additionally, researchers from the social sciences have attempted to use MAS simulation in order to inform theory building for human social systems. In this thesis the simulation approach is used to explore novel ways in which co-operation and altruism may be sustained in an MAS.

Firstly, however, a little background information is given about the areas touched upon so far.

### 1.1.2 DAI and MAS

DAI has a lineage back to traditional AI and hence has a strong methodological bias towards explicit representations of facts, goals and rules within agents (often recast within MAS as Beliefs, Desires and Intentions or BDI agents [96]). Historically DAI started from the problem of how traditional AI systems (based on, say, production rules or frames), previously developed and executed in isolation, might communicate, share knowledge, co-operate, form alliances etc. to solve particular practical problems. DAI traditionally focuses on protocols of communication and negotiation, the development of agent architectures (both practically and formally) and dedicated agent languages [167]. Interestingly, however, it is apparent that sociological issues are becoming prominent due to the difficult problems that DAI has tried to address (such as the establishment of "shared norms", "joint intentions" and the promotion of co-operative interactions). DAI workers rarely apply evolutionary techniques to address such issues. This is mainly due to the practical nature of the work. Most DAI systems and methods are designed to perform some task as optimally as possible. Agents are generally coded and debugged in the usual way. The focus is on engineering agents and the interaction protocols that form a basis for optimal global performance when composed into an MAS.

### 1.1.3 Social Science and Artificial Societies

Social scientists have argued [27], [64] that simulation models can make discourse on human societies more precise than natural language theories. The production of a concrete executable model (a computer program) and sound empirical investigation of the model can augment or replace intuitive speculation. Given a computer simulation model it is possible to show that a given set of assumptions (embodied in the program) is sufficient to produce some given behaviour - produced when the program is executed. Since assumptions can be changed by changing the program, a kind of "artificial social laboratory" [60] is then available in which a researcher can set-up various experiments to test intuitions and theories.

The construction of MAS simulations gives social scientists the ability to create artificial worlds, composed of many interacting agents, and test intuitions and theories in the artificial world. Much work in this area goes under the name of "artificial societies" [63], [52], [64]. Artificial Society work has strong methodological similarities to Artificial Life [106]. In artificial life studies, the emphasis is often on unexpected and complex life-like phenomena emerging at the macro (system) level from the complex interaction of simple subsystems. Social scientists are often interested in the formation of (potentially unexpected) macro level properties (i.e. properties at the level of the society) from given micro behaviours (i.e. individual actions). Examples of emergence in the natural world which are often cited are the flocking patterns of birds or the seemingly intelligent behaviour of an ant colony. In both cases the macro properties (flocking behaviour in birds and complex task specialisation in ants) appear to "emerge" from simple individual behaviours. There appears to be no overall plan or central controlling agent governing such "emergent" phenomena. The concept of "emergence" is discussed, in a little more detail, later (see section 1.1.5).

In much of the work presented in this thesis an attempt is made to start a dialogue between MAS and Social Science. It seems both could benefit from such a dialogue and computational models could help to ground that dialogue. However, this thesis comes from a computational modelling perspective informed by work within MAS. Where terms such as "culture", "cultural group" or "social construction" are used they are *purely related to the artificial society models presented.* Any correspondence to social science terminology relating to *real* human societies is *speculative and would almost certainly be revised in the light of inter-disciplinary dialogue.*

### 1.1.4 Artificial Life

Artificial Life models attempt to illuminate the general processes which underlie life-like phenomena [106]. Agents are generally very simple, adaptive, reactive systems (responding immediately to their inputs) and are evolved over time using some kind of Darwinian algorithm. Heavy use is made of Genetic Algorithms and Neural Network techniques. Agents are often situated within some simulated environment and allowed to move, eat, reproduce etc. Via a process of natural selection, the agents adapt over time to their environment (which includes other agents). The human observers of these models then analyse and attempt to explain the "artificial ecosystem" that has emerged. In order to understand systems of such complexity it is generally necessary for multiple runs to be performed with different pseudo-random number seeds (most models are stochastic to some degree) and adjust exogenously defined parameters for sensitivity testing. Most of the work within Artificial Life is focused on biological processes rather than the societal processes of interest within Artificial Society work. Artificial Life tends to define itself as a sub-symbolic, sub-cognitive enterprise in opposition to traditional AI. Thus although there are similarities in techniques and methodologies, Artificial Societies typically have a different focus and interpretation to Artificial Life models.

### 1.1.5 Emergence - What Does it Mean?

The word "emergence" is often used by Artificial Society and Artificial Life researchers [90], [61] to denote observed phenomena produced by their models. There is, however, no agreed or clear definition of just what constitutes emergence. The weakest and most general characterisation of the word, when applied to complex computational models, might be "any observed phenomena that was not explicitly built-in to the model specification". This kind of use is highly subjective and controversial. In some sense, it can be argued, that any behaviour of the model is by definition "built-in" to the specification. If that view is taken then, it might be more accurate to say an emergent property is "any observed phenomena that is not understandably reducible to the rules specified in the model specification". However, "understandably reducible" appears vague and subjective. Even more tentatively one might say "any observed phenomena which is not easily reducible to the model specification". However, "easily" is again rather subjective and historically dependant.

A stronger form may be characterised as "observable phenomena that may be captured by explanations (or theories) which utilise entities which are not part of the model specification". This "strong emergence" is not subjective in the sense that it makes claim that a theory can be constructed utilising entities or concepts which are not built-in to the model specification. Such entities would be of a higher level than the model specification. The claim that any phenomena observed from a model are by definition built-in to the model specification may, in some sense, be true. However that does not preclude the necessity or "use value" of theory based on higher-level entities which themselves form no part of the model specification. For example, claiming that the constitution of the United States of America is "built-in" to the laws of physics is a valid philosophical position but other kinds of theory, based on higher-level entities, are more convenient for understanding what forces produced and shaped it.

In summary, two kinds of emergence have been described here. Weak emergence is rather subjective, yet widely used, and refers to "unexpected observations". Strong emergence is not subjective. It indicates that some higher-level entity can be used to construct useful theory applicable to the given domain[1].

The identification of emergent properties requires empirical investigation of simulation models. There is currently no general analytic method of starting from some specification of a desirable emergent (macro) property and then working backwards to sets of (micro) rules that generate such properties. If there was, the conception of emergence and the engineering of MAS would be less problematic.

One empirical method of attempting to find micro rules which produce a desirable emergent macro property involves the parameterisation of the micro rules, giving a space of models. This space of models can be empirically searched, or scanned, in an attempt to locate the desirable macro properties and to characterise the space. Such a technique is applied in this thesis (see chapter 7). The techniques and tools developed, and here reported, are general and, in principle, could be applied to any model, so long as the macro property of interest can be detected automatically[2].

[1]The distinction between weak and strong emergence is made here to distinguish a purely subjective observation (weak) with an objective claim (strong).

[2]Specifically, that the macro property of interest can be detected by some algorithm. This may be far from a trivial task though a useful process in its own right.

## 1.2 Thesis Plan

In this thesis I shall report on investigations, using artificial societies, which address the problem of the emergence and maintenance of co-operation and altruism in large agent societies. The thesis consists of ten chapters. Chapter 2 introduces the problem of the origins of altruism and co-operation and relates this to current work within MAS, DAI and the biological and human sciences. Several mechanisms by which co-operation may be obtained are examined and rejected as general solutions, or explanations, for many kinds of co-operation that occur within human societies and that would be desirable within MAS. A "memetic" conception of culture is introduced and the notion of a "cultural marker", or "tag", is related to the potential for cultural group formation processes that promote co-operation and altruism. The issues raised in chapter 2 motivate the experimentation which follows in the rest of the thesis.

Chapter 3 makes a contribution to artificial society methodology by clarifying different modes of possible enquiry. Since artificial societies do not attempt to model real societies in any direct sense, work in this area can often be confusing. Chapter 3 presents a general framework which can be used to characterise and clarify experimental work which is performed with such models.

Chapter 4 gives an overview of the experimental progress (line of reasoning) for the following chapters in which artificial societies are constructed and experimentation performed. The experimental work is related back to questions raised in chapter 2 in relation to co-operation and altruism in agent societies and to the methodologies given in chapter 3. Conclusions and detailed results are not repeated in this chapter but left to their respective chapters.

Chapter 5 describes in detail an artificial society (the SwapShop) and a set of experiments conducted upon it. The SwapShop specifies a world in which simple agents decide whether to share resources with each other. A cultural learning rule is compared with a genetic learning rule in the promotion of altruism. Group processes are also observed and characterised.

Chapter 6 introduces a more complex artificial society (the StereoLab) in which agents stereotype other agents based on observable tags or "social cues". In the SwapShop society (chapter 5) agents had the ability to determine those others who were culturally identical and bias their resource sharing accordingly. It is argued that in large societies, where agents must interact with many "strangers", such an assumption is not valid. The StereoLab attempts to capture minimally a society in which interaction with strangers occurs often and agents can only use tags to distinguish between them and select appropriate actions. Many of the assumptions of the StereoLab are specified as exogenous parameters which can be varied producing a space of possible similar societies. This parameterisation enables the space of societies to be searched, locating those societies which produce the potentially desirable property of high co-operation[3].

Chapter 7 addresses the issue of how a parameter space (such as that specified in chapter 6) of an artificial society might tractably be explored and characterised. Given a large high dimensional space (as the StereoLab parameter space is) various methods of exploration are described. The application of a decision tree induction algorithm (C4.5) and hill-climbing, combined with cluster analysis, are selected as applicable techniques. A link between different parameter space exploration methods and the methodologies given in chapter 3 is also made.

Chapter 8 presents detailed experimentation with the StereoLab society (described

[3]Here "desirable" is meant in the context of the optimality of MAS measured in the achievement of individual agent goals. See section 2.1 in chapter 2 for a discussion of this.

in chapter 6) through the application of the parameter exploration techniques described in chapter 7. In addition to characterising the levels of co-operation found within given regions of the parameter space, individual simulation runs are selected from identified regions and examined in-depth. From the analysis of such individual runs, explanations are offered which relate the co-operation forming processes occurring within the regions to the parameter ranges which characterise them. More simply put, explanations are advanced which answer the question of why a given region of the space gives rise to a particular level of co-operation.

The chapter concludes with the identification of an interesting region in which apparent group formation processes based on tags promote high co-operation. In order to isolate and explore this process in more detail a further artificial society is presented in chapter 9 (TagWorldII) which captures, minimally, this tag based group formation process. The greatly reduced complexity of the TagWorldII society allows for an in-depth examination of the underlying process, including visualisation of group formation processes over time.

Chapter 10 summarises the main contributions of the thesis and relates results obtained to MAS and human societies. Possible future work is also outlined. Of particular interest are the results obtained in chapter 9. Here a new form of "cultural group selection" is demonstrated. It is based on a radically different conception of what constitutes a "cultural group" compared to previous models. The possible applications this form of cultural group selection in MAS engineering and increased understanding of human societies are discussed.