Chapter 10

Conclusions and Future Work

In this thesis the problem of modelling sustained pro-social behaviours in multi-agent systems has been examined. Work within DAI has tended to adopt an engineering perspective to address such questions, but here a cultural evolutionary perspective has been applied. Empirical investigations were performed on three different multi-agent simulation models (artificial societies). In each, the importance of group processes was emphasised in the promotion of pro-social behaviours. In the following sections the conclusions and deliverables of the thesis are outlined (section 10.1 below) as are directions of future work (section 10.3 below). In section 10.2 more general implications and applications are discussed.

10.1 Conclusions / Deliverables

Below is a summary of the main conclusions and deliverables of the thesis:

- A general methodological framework applicable to empirical investigation of artificial society computer models was proposed (chapter 3). It is argued that most work with artificial societies can be cast within this framework.

- It has been empirically demonstrated that agents employing simple cultural learning rules can evolve altruistic behaviours within an in-group of culturally similar agents (chapter 5). Although the spreading of cultural units (memes) can be seen as selfish, agents behave in an altruistic way towards their in-group of "cultural kin".

- A general purpose software tool has been constructed which allows for simulation runs to be scheduled, executed and results collated, in parallel over a set of computers on a network (chapter 7 and appendix A).

- Methods making use of machine learning (the C4.5 algorithm), local search (hill-climbing) and cluster analysis have been applied to an artificial society with many exogenously defined parameters in order to locate regions of interest (chapters 7 and 8). The methods used are general and could be applied to other simulation models.

- A novel co-operation forming process, based on identifiable cultural markers or "tags", was identified and described (chapters 8 and 9). This novel process was identified under both agent satisficing and agent optimising assumptions. It is argued that this may be viewed as a novel form of "cultural group selection" (see section 10.2.2 below). This mechanism may have applications within DAI (see section 10.2.3 below).

10.2 Discussion

The work in this thesis has centred around the notion of co-operation within agent societies in which simple forms of cultural evolution were modelled. Culture has been represented within the model as units (memes) which are copied between agents during cultural interactions.

10.2.1 General Contribution

The major element of the thesis was the construction of an artificial society with a large number of variable exogenously specified parameters and the search within the parameter space for novel co-operation promoting regions. Through this method a form of exploratory "theory construction" was applied in the artificial domain. This is a kind of "reverse engineering", starting with a desired macro (population level) outcome then searching for some set of micro-rules (agent behaviours) which produce that desired behaviour. The results presented here demonstrate that such an approach *can* produce results. The method could be applied to engineering problems within DAI when societies are required with certain properties but there is no theory which indicates the micro properties of the agents that are required. In this way the method presented could contribute to agent based computer science.

The area of "memetics" is widely regarded as a controversial "theory" of cultural evolution [77], [72], [4]. Specifically, the question remains as to just what constitutes "meme theory" and how such theory might be developed and communicated. The work presented here attempts to address these problems by using agent based models as a way of formalising cultural phenomena. This allows for concrete analysis, comparison and reproduction. Terms such as "culture", "meme" and "agent" are made concrete in the assumptions of the computer model. Although the use of such terms in this thesis are *provisional* and *relate only to the artificial societies presented* it is hoped that a dialogue with researchers within the social sciences can be opened. Debates over the meanings and principled use of such terms would hopefully be clarified.

10.2.2 Cultural Group Selection

The work in this thesis culminates with a proposed novel form of co-operation based on a group process. It is argued here that this may be viewed as a new kind of "cultural group selection". The concept of cultural group selection is a controversial area of debate within the anthropological community [150]. Group selection within the biological sciences is currently a non-orthodox theory but is proposed by some [165]. The work presented here demonstrates that group functional behaviours can emerge in societies of agents even when this is to the detriment of individuals within those groups.

The results obtained with the TagWorldII society presented in chapter 9 show that even egotistical agents evolving as units of evolution under the assumptions of the replicator dynamics can produce high-levels of group functional behaviour (co-operation is selected for in the one-shot Prisoner's Dilemma).

Soltis et al [150] compared anthropological data against the cultural group selection hypothesis and concluded that the rate of extinction of groups was too low to be effective in producing recent group-functional behaviours in modern societies. However, in their work they view cultural groups as follows:

> "We define a group as a territorial population that can conduct warfare as a unit. An extinction is said to occur when (1) all members of a group are killed or (2) members of a group are assimilated into another group either wholly or in part. When a group is routed from its territory but remains intact as a social unit (or its fate is unknown), then a forced migration, not an extinction, is said to have occurred." [150].

This "common sense" definition of a cultural group is very different from that which is used within the TagWorldII model. In the TagWorldII model physical space is irrelevant. It is not modelled. Groups inhabit "tag space". The tags (social markers or labels) demarcate cultural groups. Tags are culturally acquired and can be changed by

mutation or imitation[1]. A cultural group is defined as a set of agents sharing the same tags. Agents limit their game interactions to those within their own group. Since tags can be changed very quickly, groups come and go very quickly. They are highly ephemeral and permeable - individuals may enter and leave easily. It is argued that this view of cultural groups is more applicable to large modern social systems (e.g. the post-industrial metropolis) in which cultural identity is changeable and ephemeral - a fashion statement or brand image. It is the very fact that such cultural groups are precarious and insubstantial that allows them to become a force for selection since extinction rates can be very high. Extinction in this view is not related to agent death but simply to the movement of agents in tag space (selecting different tags). The TagWorldII model does not model physical space or agent death. The model attempts to show how groups can form and evolve intra-generationally through cultural processes. The asexual selection and reproduction process in the TagWorldII model is meant to capture the spread of behaviours via an imitative process, not agent birth and death.

This different view of cultural groups raises the question of how an empirical analysis might be applied to real societies in order verify this cultural group selection hypothesis. This appears problematic. Firstly, individuals will be members of many (possibly overlapping) cultural groups. It would also seem unlikely that membership of a specific group could be easily reducable to the presence or absence of a few observable characteristics. Even if such a group could be identified how could it be measured over time? Perhaps more problematically, how could such groups be identified retrospectively? Anthropological evidence can document the extinction of spatially located and bounded groups over time but does such evidence exist for the ephemeral cultural groupings in modern societies realised within

[1]TagWorldII does not model the imitation process directly. It is assumed that selection based on utility and asexual reproduction approximates to it (when agents imitate those who have higher utility than themselves).

TagWorldII?

These problems suggest that empirical verification of the cultural group processes found within TagWorldII would require significant social scientific work. However, perhaps one way forward is to study the new kinds of social space emerging through the internet. It is well documented that the internet facilities the rapid creation (and dissolution) of many "electronic communities" in which social interaction and even economic interaction takes place. It would seem that physical space is irrelevant in such a setting. Also individuals can only judge each other by a limited set of cues. Perhaps this could be a basis for an empirical investigation[2].

Alternatively tags could be applied to existing artificial society models which capture significant aspects of human societies. For example, in Cecconi and Parisi's [24] model of the non-emergence of "social survival strategies", in groups of agents, the application of tags could be used to evolve such social strategies[3].

10.2.3 Group Selection in DAI

Kalenka & Jennings [100] applied and compared various forms of "social rationality" in a simulated warehouse unloading scenario. Individual agents have responsibility for unloading individual bays but may help others if they have free time between unloading. In these experiments agents were set with various interaction behaviours ranging from selfish to altruistic. The results demonstrated that the agents inclined to behave altruistically

[2]One could imagine that "spamming" (i.e. unsolicitied advertising) and "flaming" (i.e. abusive and offensive comments) of special interest newsgroups could be seen as a form of "defection" whereas the posting of high quality information as a form of co-operation. When the overall quality of information posted to a newsgroup is low we might expect it to "dissolve" (i.e. cease to be posted to or read). The feasibility of this kind of investigation is demonstrated by Best [12] in which newsgroup postings were analysed in a memetic context.

[3]Their model is essentially an n-person Prisoner's Dilemma. However, they model *the underlying process*, in a food foraging and sharing scenario, by which such a dilemma emerges.

tended to increase the performance of the system overall[4].

Future work could examine the application of tag based cultural group selection to the same scenario. Such an application would involve sets of competing groups (or teams) of agents. Agents would only be able to elicit or give help to those within their team but would be able to move between teams at given intervals. The decision made by an agent of when to help another group member would constitute the evolving "strategy" in such a scenario.

Although intuitively it seems that such an evolutionary process should produce desirable results the more realistic nature of the warehouse scenario would raise interesting issues which are not present in the very abstract TagWorldII model. By attempting to apply cultural group selection in this more concrete form progress could be made towards a practical application of the technique within multi-agent systems engineering.

10.2.4 Critical Review

The "theory construction" method used in the thesis highlighted a number of problems. The problem of exploring a large parameter space was tackled by automatically categorising individual simulation runs based on a single output measure (co-operation). In order to locate more complex emergent properties, more complex measures would be required. It seems hard to envisage how such complex measures of emergent properties can be specified algorithmically and *a priori*. Previous work with the StereoLab [71] which attempted to find regions in the parameter space with particular group properties was only marginally successful - the measures became increasingly complex and interpretation of results became non-trivial. Even when analysis was focused on a simple measure (co-

[4]This summary is a simplified picture. Interestingly it was found that in some circumstances too much altruism could hinder performance since agents put too little weight on their own goals and this had a detrimental effect on system level performance.

operation) questions still remained which were hard to answer *a priori*. For example, in the work presented in chapter 8 each simulation run was executed to 100 cycles. The value was essentially arbitrary, chosen because it made the large number of runs required computationally tractable (see future work below). Co-operation was measured over the last cycle, but this could just as easily have been an average of all cycles. Particularly relevant here is the stochastic nature of the model (and many agent based models have some stochastic elements). A single run from a given point in the parameter space may produce many different trajectories depending on the pseudo-random seed used. Should a single run count as evidence that particular phenomena is likely for some given set of parameters? If multiple runs are used with different pseudo-random seeds then how many should be used? It seems hard to answer such questions *a priori* without some notion of how the particular model behaves. But the idea behind "theory construction" suggests that we have few *a priori* assumptions. If we know what the model is going to do then in some sense we already have a theory.

Although the work presented here *did* produce a novel theory of group based co-operation (as evidenced in chapter 9) the method employed did not *guarantee* this. A less risky approach would require a more focused set of experimental objectives but this would require more *a priori* theorising. In the work presented in chapter 8 both C4.5 and k-means cluster analysis were used to find regions in the parameter space. Although both produced insight, cluster analysis gave the more interesting results - however, C4.5 was not applied to the "extended space" which produced the results that lead to "cultural group selection" theory.

10.3 Future Work

10.3.1 Short Term Future Work

Given the work presented in this thesis, below is a summary of possible *immediate* future work:

- Modify the TagWorldII artificial society (chapter 9) such that agents may evolve the level of tag biasing they apply. The bit string representing each agent would be augmented with an individual "refusal" value. This value would indicate the number of refusals the agent would be allowed to make before forced game interaction. It would be of interest to know if high refusal values would evolve. If this were the case, the current *assumption* of a high refusal value could be justified by reference to an evolutionary process.

- Modify and apply existing analytical patch based models of altruism developed within sociobiology to the topology of the "tag space" and the group dynamics within the TagWorldII. Such an analysis would require estimates of group formation, group extinction and group migration rates. Initially these values could be estimated empirically from the TagWorldII model. If an analytical patch based model could be produced, which captured the TagWorldII dynamics, then predictions could be made as to the range of parameters which would support co-operation between agents. Those predictions could be tested empirically against the TagWorldII model.

- Port, test and improve the SampTool software such that it could be easily deployed on most popular computing platforms and release it for use by other researchers.

- Apply the C4.5 algorithm to the "extended space" (to which k-means cluster analysis

was applied in chapter 8) for purposes of comparison. Would C4.5 find the regions of interest that lead to the "group selection" theory?

- It would be of interest to try running the StereoLab simulations for longer numbers of cycles. The current limit of 100 cycles is arbitrary. Would longer runs produce significantly different results? Speculatively it could be hypothesised that longer runs would produce more co-operation, since agents have more time to co-ordinate their actions.

10.3.2 Long Term Future Work

More general directions for future work leading from this thesis:

- Application of "cultural group selection" to the simulation of a more complex task oriented DAI scenario (e.g. the robot unloading scenario as described in section 10.2.3 above). Results could be compared to individual and socially rational agents.

- Application of "cultural group selection" to Cecconi and Parisi's scenario [24]. It would seem that the tag based process described here may be sufficient to evolve the "social survival strategies" which were posed by the authors as crucial in understanding the emergence of human societies.

- The Application and comparison of other machine learning techniques (e.g. BACON [105]) in the understanding of results derived from simulation models. Specifically to explore what kinds of induction methods work best in what kinds of situations. More specifically, to ascertain if there is anything unique to agent based simulation that can be exploited by such inducers.

- To explore methods by which the various methodologies detailed in chapter 3 may

be semi-automated (e.g. in a similar way to the Weaver system [81]) such that a re-usable and general "simulators laboratory" is produced.