Models of Multi-agent Decision Making.

Julian Zappala

School of Computer Science

University of Nottingham

Thesis submitted to The University of Nottingham
for the degree of

Doctor of Philosophy

July 2014

Abstract

In this thesis we formalise and study computational aspects of group decision making for rational, self-interested agents. Specifically, we are interested in systems where agents reach consensus according to endogenous thresholds. Natural groups have been shown to make collective decisions according to threshold-mediated behaviours. An individual will commit to some collective endeavour only if the number of others having already committed exceeds their threshold. Consensus is reached only where all individuals express commitment.

We present a family of models that describe fundamental aspects of cooperative behaviour in multi-agent systems. These include: coalition formation, participation in joint actions and the achievement of individuals' goals over time. We associate novel solution concepts with our models and present results concerning the computational complexity of several natural decision problems arising from these.

We demonstrate potential applications of our work by modelling a group decision problem common to many cohesive groups: establishing the location of the group. Using model checking tools we compute the effects of agents' thresholds upon outcomes. We consider our results within an appropriate research context.

Acknowledgements

Foremost, I gratefully acknowledge the assistance of Dr. Brian Logan and Dr. Natasha Alechina. Their kind encouragement and guidance whilst supervising this research has been, and continues to be, invaluable. Undoubtedly, without their patience, inspiration and unfathomable knowledge and experience this thesis would not have been possible.

Thanks also to the academic, technical and administrative teams at the School of Computer Science, University of Nottingham, who provide such a rich environment in which to study and work. During the preparation of this thesis I have had the pleasure of meeting countless individuals, staff and students alike, who have contributed towards its completion. In particular I would like to thank Dr. Tony Pridmore, Dr. Peter Blanchfield, Dr. Peer Olaf Siebers and Dr. Dario Landa-Silva for their advice and support.

I am indebted to Dr. David Parker (School of Computer Science, University of Birmingham), lead developer of the PRISM model checking project. His responsiveness and willingness to extend this tool greatly accelerated progress of the applied work contained in this thesis. Thanks also to Dr. Andrew King (Swansea University's College of Science) for his generosity in sharing a biological perspective on this work.

Lastly, I wish to thank my family and friends, including Susan, Mary, Carla, Gerald and Derek, for their help and cooperation throughout this endeavour.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Introduction

In distributed artificial intelligence a multi-agent system is comprised of a number of discrete yet interacting computational processes known as agents [Weiss, 2000]. Although the agent community lacks a unified view of precisely what qualifies as an agent [Siebers and Aickelin, 2008] common views of agency include notions of autonomy of action, social-ability [Wooldridge and Jennings, 1995] and goal orientation [Franklin and Graesser, 1996]. If the potential of multi-agent systems is to be realised it is essential that agents are able to cooperate and coordinate their behaviours [Kraus, 1997].

For certain classes of multi-agent system, where agents are inherently cooperative, coordination may not necessarily present significant challenges. Indeed, it is often the coordination of activity, typically about some common goal, which distinguishes a *team* of agents from a collection of *individual* agents. In this thesis we largely concern ourselves with systems of the latter type.

The need for coordinated collective behaviour is present even where agents are not naturally inclined towards cooperation. There are many reasons why agents may wish to, or indeed have to cooperate, for example, where resource is constrained or otherwise in contention [Youssefmir and Huberman, 1995], where agents have differing abilities [Grosz and Sidner, 1990], or where they possess differing information [Pigozzi, 2006]. Competitive, self interested, agents may benefit from cooperative behaviour where this is consistent with individual rationality, for example, where cooperation increases their individual utility. However, inducing cooperative behaviour amongst agents with heterogeneous and possibly conflicting goals is not straightforward.

Typically coordinated behaviour requires that the agents can reach agreements regarding, for example, which agents should cooperate, and what actions those agents should perform. The predominant models of coordination within multi-agent systems have followed from theories founded in economic or political settings. We draw inspiration from elsewhere. In this thesis our interest is in a coordination mechanism which is prevalent in the natural world.

Many phenomena displayed by biological/social groups can be explained in terms of individual responses conditioned upon thresholds. Typically, each individual's threshold represents a minimum number of others which must first commit to some joint endeavour prior to that individual following suit. Such *threshold models* describe collective activities in a variety of species from the simplest organisms [Jacob et al., 2004] to the higher primates, including ourselves [Schelling, 1969].

Empirical research has shown that many biological systems employ threshold mechanisms as a means of making agreements, or more generally, reaching consensus. For example, social insects including ants and bees are known to use a threshold mechanism termed *quorum sensing/response* when deciding between alternative nesting sites. The ubiquity of quorum decision making in nature suggests that this is an efficient, effective and stable mechanism through which a group's activities can be coordinated. Theoretical models support this view, predicting that where thresholds are adaptive, decisions can not only be optimal [List, 2004] but also offer a trade-off between speed and accuracy [Pratt and Sumpter, 2006].

The practice of constructing artificial systems which are in some way inspired or otherwise motivated by natural systems is well established in the field of computer science. Such work has yielded results in areas as diverse as robotic control [Beer et al., 1997], security [Greensmith and Aickelin, 2008], and optimisation [Kennedy et al., 1995]. Indeed, the potential of quorum-based decision making has not escaped the computer science community.

Applications for threshold mediated behaviours have already been proposed in diverse areas such as mobile networking Peysakhov et al. [2006] and computer security Vogt et al. [2007]. However, despite this interest, relatively little attention has been given to formal models and computational properties of threshold behaviours in multi-agent systems. We shall consider such matters here.

1.2 Background and Motivation

This work was inspired by an agent-based model of group decision making in baboons [Sellers et al., 2007] in which a simulation of group decision making in a troop of chacma baboons (*Papio hamadryas ursinus)* is created and investigated. The simulation composes two models; the first representing the environment, and the second representing an agent. These models are briefly outlined below.

The environment model is based on field data collected from the baboons' natural habitat, the De Hoop Nature Reserve in South Africa. The environment is modelled as a grid of 660 cells representing an area of 5.4km $\times$ 8.4km. Each cell represents an area of 200m $\times$ 200m map and contains a mixture of habitat types and food sources. Cells may also include one or more 'special features' including water sources, sleeping sites, and refuges.

Each baboon is modelled as an agent with physical parameters based on well-known baboon physiology. Each agent maintains an individual 'desire score' for activities key to the individual's survival; these include scores for drinking, eating and time spent socialising. These scores serve to bias the agent's choice of preferred activity. The ability to engage in a preferred activity is constrained by location since, for example, the drinking activity requires that water is present.

At each time-step of the simulation the agents must make a group decision concerning the location of the group. Each agent's preference for the group's location is informed by their choice of preferred activity. For example, 'thirsty' agents may wish to move towards a cell containing water, whereas, 'tired' agents may wish to move to a cell suitable for sleeping.

The problem of choosing a location for the group is challenging since the group operates under the constraint that they must not separate; the survival chances of a lone baboon are not great. Since the agents may prefer different locations (based upon their desire scores) conflicts concerning the group's location can arise. Such conflicts are resolved using a democratic voting process governed by a variable threshold. At each time-step of the simulation the group vote over possible outcomes including whether or not the group should move, and if so, in what direction. Outcomes receiving a number of votes in excess of the threshold are "winning"; ties are resolved using non-determinism.

A significant result of Sellers et al. [2007] is that the choice of voting threshold is the most important determinant of both the fitness of each agent and the behaviour of the group as a whole. Thresholds of between 50% and 65% provided agents with the greatest chances of survival; outside of this narrow band the chances of an agent failing to meet its survival targets, measured in terms of food/water intake and time spent in social engagement, increased drastically.

The simulation of Sellers et al. [2007] describes a group, or *coalition* of agents that must make a collective decision. The outcome of this decision is a collective act: to remain at the present cell or to move, as a group, to an alternative cell. Participation in this collective act has implications for the individual goals that each agent can achieve. For example, those agents which are thirsty may achieve the goal 'to drink' (by moving to a cell containing a water source) whereas those agents which are tired do not achieve the goal 'to rest' since they move together with the group.

Although there are anecdotal accounts suggesting that natural primate groups may engage in voting-like activities there is little evidence to suggest that primates, such as baboons, possess the cognitive and social processes necessary for effective group decisions to be consistently made in this way [Sellers et al., 2007]. Moreover, modelling collective decisions through the use of an exogenous and prescriptive mechanism (the voting procedure) seems to disregard the *individual* emphasis associated with agency. This observation was pivotal in motivating interest in creating models of group decision making and collective action for agents where decision outcomes are emergent as a function of endogenous thresholds, associated with each individual.

1.3 Objectives and Contributions

In this thesis we develop formal apparatus to describe threshold mediated decision making. In particular our interest is in consensus decisions, those to which all participants agree, as they pertain to multi-agent systems. Therefore, we aim to develop models describing group decisions for autonomous entities capable of acting, individually or jointly in pursuit of possibly heterogeneous goals and over time.

Aspects of our models have previously been studied, often in isolation and in disparate research strands spanning the disciplines of computer science, sociology and biology. Research in these fields has yielded models suitable for describing only specific problems in certain domains, or only a subset of the qualities typically associated with rational, self interested individuals. Consequently neither the extant models themselves or the techniques by which they have been studied are easily transferable to more general problems within the domain of interest, or similar problems in non-related domains. In this work our objectives include the creation of models that are:

- domain neutral,
- framed in terms consistent with common notions of agency,
- open to study using established methods.

The potential of artificial systems comprised of components utilising threshold mediated behaviours is *already* being explored. Given this, it is extremely interesting that the challenges of computation in systems of this kind are not well understood. We include in the objectives of this work:

- identification of fundamental questions concerning computation of consensus outcomes in threshold mediated systems,
- definition of decision problems corresponding to such questions,
- examination of the computational complexity of these problems.

It is our fervent belief that the models presented in this thesis can find application in a variety of research fields including computer science, sociology and biology. To this end the objectives of this research include:

- demonstration of potential research applications for our models through the creation of a concrete model instance,
- utilisation of our models to study known research questions apparent in literature concerning consensus decision making in natural groups,
- comparison of results obtained from our models to those found in the relevant literature,
- demonstration of the models' potential to enable novel research.

In satisfying our research objectives this thesis makes contributions to a variety of fields. It has been argued that in any distributed system which requires processes to coordinate that system must provide a mechanism through which those processes can achieve consensus [Turek and Shasha, 1992]. Prior to this work consensus via threshold mediated mechanisms has been overlooked by those interested in general modelling of distributed computation. Our results regarding the tractability of problems concerned with reaching consensus in this way are therefore valuable to researchers interested in coordination within any distributed computational environment.

Our particular interest is in models for multi-agent systems. Research in this field is most commonly concerned with the behaviour of agents that are in some-way rational. In the course of this thesis we make several contributions to the understanding of rationality within the context of threshold mediated consensus. We study notions of consensus that are both individually and collectively rational and associate a number of novel solution concepts with these. Furthermore we propose a series of postulates that constrain the selection of possible threshold values to those which are rational under a representation of the agents' goals.

At the time of writing the majority of research concerning consensus in threshold mediated decision making has taken place in the fields of biology and sociology. In contrast, we study this phenomenon from a computational perspective using tools and techniques particular to this domain. Our work demonstrates new methodologies for investigating threshold mediated consensus in biological and sociological systems. Ours is the first work to study consensus decision making in natural groups using formal verification techniques.

1.4 Thesis Structure

In this thesis we develop and characterise a series of models. The structure of this thesis and the development of our models are consonant with our objective to create models that intuitively capture common notions of agency. Our first model considers only the problem of coalition formation for agents whose behaviours are mediated by threshold responses. This model captures the social nature of agents but is otherwise quite abstract. We proceed to extend this model such that it encompasses more 'real-world' situations where agents may jointly perform individual actions and then, through these, achieve their individual goals. The development of these models occurs within chapters three to five.

Our final model is presented in chapter six where we develop a temporal model capable of describing agents' behaviours over time. We embed the previous model within a finite state transition system where each state represents an instance of that model and each transition represents one or more possible joint actions the agents may perform. In chapter seven we turn our attention to the utility of our work by demonstrating a potential application for the final model. Returning this work to its origins, the model presented in chapter six is grounded in an idealised group decision making scenario inspired by the work of Sellers et al. [2007].

For each model we analyse the computational complexity of decision problems concerning the verification, existence and non-existence of consensus outcomes for the agents. By developing our models incrementally we are able to study the contribution that each aspect of agency makes to the computational complexity of these fundamental problems. This approach reveals some surprising results. In particular, is not necessarily the case that a more expressive model leads to an increase in complexity for comparable decision problems. We find that by exploiting new structures, as they are introduced, certain problems which appear computationally intractable in early models are subsequently found to be tractable.

Specifically, the remainder of this thesis is organised as follows:

- **Chapter 2 - Related Work**: We survey the extant approaches to coordination, cooperation and agreement for multi-agent systems. In particular we focus on models drawing from the fields of game-theory and social choice. We also introduce relevant concepts from the biological and sociological literature concerning threshold models and quorum-based consensus.

- **Chapter 3 - Consensus Games (CGs)**: We present a model of coalition formation for multi-agent systems via threshold-mediated behaviours together with notions of stability for this model. We examine a number of natural decision problems arising from these solution concepts. We then consider the implications of this model for distributed settings where agents may have incomplete information.

- **Chapter 4 - Consensus Action Games (CAGs)**: The model is extended to encompass coalitions that form in order to perform some joint action. We consider the implication of this extension for the previously defined notions of stability. We consider the complexity of decision problems analogous to those previously considered and draw comparisons between these results and those for CGs.

- **Chapter 5 - Consensus Action Games with Goals (CAG-Gs)**: The model is extended to encompass coalitions that form in order to perform a joint action that achieves one or more individual goals. We consider the rational selection of threshold values where agents have heterogeneous goals and heterogeneous orderings over these goals. We continue our analysis of the computational complexity of decision problems for this model and draw comparisons between these results and those for CAGs

- **Chapter 6 - Temporal Consensus Action Games with Goals (T-CAG-Gs)**: The model is extended to encompass coalition formation, joint action and goal achievement, repeatedly, over time. We formalise temporal solution concepts for this model. We formulate and consider the computational complexity of decision problems for this model.

- **Chapter 7 - Demonstration**: We apply the T-CAG-G model of the previous chapter to a concrete scenario concerning consensus decision making in natural groups. We use model checking techniques to formally verify properties of the model, including the existence of stable solutions. We measure the effects of agents' thresholds upon the likelihood of consensus and positive outcomes. We compare our results with those found in the associated literature.

- **Chapter 8 - Conclusion**: We summarise the main contributions of this thesis and suggest possible avenues for future research. Lastly, we give details of the dissemination of this work.

Chapter 2

Related Work

2.1 Cooperation and Coordination in Multi-agent Systems

The need for cooperation and coordination amongst agents is generally accepted to be amongst the central challenges for designers of multi-agent systems. The study of such problems is therefore prevalent in multi-agent systems literature; many modes of cooperation and coordination have been considered. Examples include the problem of task allocation, or "Which agents will complete which tasks?" [Bonabeau et al., 1997; Sandholm, 1998; Shehory and Kraus, 1998]; the problem of resource allocation, or "Which agents will receive which resource(s)?" [Bredin et al., 2000; Chevaleyre et al., 2006] and the problem of forming cooperative alliances, or "Which agents should work together?" [Klusch and Gerber, 2002; Wooldridge and Dunne, 2004].

One of the earliest approaches to coordination in multi-agent systems is the contract net protocol [Smith, 1980]. The protocol enables agents to work together to achieve tasks in situations where, for example, agents' abilities are heterogeneous and no individual agent has the necessary abilities to single-handedly complete a task. During enactment of the protocol an agent, or *manager*, recognises a task to which it is not suited, the existence of this task is then *announced* to the other agents. Such announcements may be general (broadcast) or more targeted dependent on the knowledge the manager possesses regarding the capabilities of other agents. When an agent becomes aware of a task to which it is suited that agent may make a *bid* for the task. Bids are evaluated by the manager agent which will award the task to one (or more) of the bidding agents. These agents become *contractors* for the task. The contractor then *expedites* the task; this may include sub-contracting and so managing other agents.

Managers may request updates on the progress/status of tasks and contractors report to managers once a task is complete.

The contract net has been widely implemented and is one of the most studied approaches to distributed problem solving [Wooldridge, 2005, p. 197]. However, there are some significant limitations associated with this approach. Most notably the contract net assumes that agents are *benevolent* in nature and therefore the agents' goals will never be in conflict. As Nwana et al. [1996] observe, the presence of conflict is 'one key reason why coordination is needed in the first place'.

The contract net protocol was inspired by the tendering processes used by many commercial organisations. Other approaches to cooperation and coordination for multi-agent systems are also metaphors for real-world solutions to these problems. Much of our everyday behaviour is governed by social norms or laws; for example, in the United Kingdom, vehicles drive on the left hand side of the road. This rule brings about cooperation and coordination amongst road users and (to a large extent) ensures that the road system as a whole operates smoothly. As with human societies, such constructs can serve to establish expected and acceptable behaviours in agent societies.

There has been considerable interest in the specification and development of *normative* multi-agent systems. Elakehal and Padget [2012] identify three ways in which this approach may be applied. The first is through design-time constraints where desirable behaviours are hard-coded into the agents. In the second method protocol-level constraints are used to enforce conventions; this has become known as *regimented* norms. Thirdly, *regulated* norms can be used to encourage agents to act in a specified manner.

The principal difference between regimentation and regulation is that in regulated systems it is not obligatory that agents comply with the system's norms. Hence, where norms are regulatory, agents may act in violation of the norms. Consequently the designers of regulated normative systems are often interested in designing systems that incentivise agents towards norm compliance. Typically this involves invoking some form of *sanction* against agents whose behaviour is non-compliant. Possible sanctions include financial penalties, restriction or denial of access to the system for some time period, or a reduction of the trust/reputation associated with the agent [Vázquez-Salceda et al., 2004].

Whereas the contract net protocol assumes that conflicts between agents do not exist, normative systems attempt to reduce or eliminate potential conflicts. Other approaches to cooperation and coordination have focused on the resolution of conflicts where they occur. One such technique is multi-agent negotiation.

In multi-agent negotiation [Rosenschein and Zlotkin, 1994], as in human negotiation, agents aim to find a compromise position - one to which all agents involved in the negotiation can subscribe. Typically multi-agent negotiations proceed in a series of rounds. During each round each agent makes a *proposal* regarding their preferred outcome to the negotiation. If there is no proposal that meets general agreement the agents modify their proposals for the next round. Negotiation continues until a suitable compromise agreement is found, or some other termination criteria is met.

In Beer et al. [1999], Jennings identifies three main elements of research into multi-agent negotiation. Negotiation *protocols* describe the process by which negotiations are undertaken. This includes rules concerning the permissible participants and actions during the negotiation, the possible states of the negotiation and significant events of the negotiation procedure. The negotiation *objects*, sometimes referred to as the negotiation set, are the range of possible outcomes for the negotiation. Lastly, the *reasoning models* of the agents describe the strategies employed by the agents when selecting their proposals.

Normative and negotiation based approaches to cooperation and coordination in multi-agent systems consider actions by rational, self-interested agents. Both approaches require that agents evaluate the costs and benefits associated with their behaviours. For example, in regulated normative systems agents must evaluate the cost (sanction) associated with violating a norm, against possible benefits that this may bring about. Recently there has been considerable interest in viewing such strategic interactions from a game theoretic perspective. We shall consider this approach in the following section.

2.2 Coalition Formation in Multi-agent Systems

A crucial issue for agents when faced with choosing between alternative cooperative options is determining which coalitions of agents should form [Wooldridge et al., 2007]. The problem of coalition formation, or “Which agents should form cooperative alliances?” is fundamental to cooperation between agents.

The study of coalition formation has its origins in the game theoretic models founded in the field of economics. In non-cooperative games [Nash, 1951] each player may play one of a number of strategies. The reward, or pay-off a player receives is dependent on the strategies played by the others. Such models are the predominant paradigm for studying strategic interactions between multiple individuals.

Central to the analysis of these games are the properties of stability and rationality. An outcome is said to be in Nash equilibrium if no individual can benefit from individually deviating from their current strategy; in this situation each individual is playing their best response to the strategies selected by the others. Such equilibria are stable in the sense that they are *individually* rational; however, Nash Equilibria may not be *collectively* rational. That is, there may be an alternative outcome where *all* players receive a better reward.

This is highlighted by the prisoner's dilemma; a game where two criminals are suspected of a crime. The suspects are interviewed separately; each is offered freedom in return for their testimony (defection) against the other. The terms (pay-offs) are: both remain silent and each receives a short custodial sentence, both turn informant and receive a medium sentence, where only one defects the other receives a long sentence. The Nash equilibrium for this game corresponds to both suspects defecting. The best collective strategy is that both remain silent, however each fears that the other will defect. The suspects' difficulty is that they have no way to agree this plan.

The model of cooperative games [Von Neumann et al., 1947] describes players that seek to form coalitions of 'like-minded' individuals. Players within a coalition commit to a binding agreement regarding their behaviours on the understanding that their collective action will yield a collective reward or pay-off which is then, according to some fair process, shared amongst members of the coalition.

A desirable quality for any coalition to form is that the coalition can be said to be stable. That is, once formed, the coalition will not fall apart. A coalition might collapse where, for example, certain members of that coalition could increase their reward by breaking away; these members have an incentive to deviate from the coalition. Coalitions that are immune from such defections are said to be in the *core* of the game.

The traditional treatment of cooperative games embodies the assumptions that the reward, or utility, due to the coalition is in some way commonly accepted by, divisible amongst and transferable between players within that coalition. These are reasonable assumptions where the setting for the game is economic and so utility is a monetary quantity. However, they do not necessarily translate well into more general settings where a quantitative worth cannot be ascribed to a coalition, or where there is no globally accepted form of utility. Such issues have led to an interest in alternative models for coalition formation, such as the model of non-transferable utility games, or games without side-payments [Aumann, 1961].

One example of cooperative games with non-transferable utility, framed specifically in terms of multi-agent systems is cooperative Boolean games [Dunne et al., 2008]. In these games agents have a single goal, expressed as a formula in propositional logic over a set of Boolean variables. Each agent may influence the truth values of disjoint subsets of these variables through their choice of action. Agents are motivated towards cooperative action since the achievement of their goal expression may require that values of variables over which they have no control are altered. A cost is associated with each action; whilst agents are interested in achieving their goal they are also interested in minimising costs. Costs incurred by agents are dependent upon which coalition the agent participates in, hence agents form preferences over coalitions (this resembles hedonic games, e.g., [Elkind and Wooldridge, 2009]). Dunne et al. [2008] define a number of solution concepts for Boolean games and present results concerning the computational complexity of ascertaining whether some outcome is stable.

Dispensing entirely with the notion of utility, Wooldridge et al. [2007] propose a class of games where agents are not motivated to either minimise costs or maximise rewards. Rather, they are interested in achieving one or more goals. They term these games Qualitative Coalitional Games (QCG).

In QCGs each agent, in a finite set of agents (Ag), seeks to achieve one or more goals drawn from a common, finite set (G); agents are considered to be 'satisfied' where at least one of their goals is achieved. More formally, a QCG Γ may be represented as the (n+3) tuple $\Gamma = \langle G, Ag, G_1 \ldots G_n, V \rangle$ where $G_i \subseteq G$ represents each agent's $i \in Ag$ set of goals and $V : 2^{Ag} \rightarrow 2^{2^G}$ is the characteristic function of the game mapping each possible coalition of agents to the set(s) of goals that coalition can achieve. QCGs describe a wide variety of coalitional scenarios including, for example, participation in voting procedures [Wooldridge et al., 2007].

Stability in QCGs is characterised by a notion similar to that of the core in more traditional coalitional games. Specifically, a goal set is in the core (and hence stable) for a coalition where every member of the coalition can achieve at least one of their goals (every member will be satisfied), the characteristic function returns the goal set for the coalition (the goal set is feasible for the coalition) and the removal of any member of the coalition renders the goal set infeasible for the other members of the coalition (the coalition is minimal). Wooldridge et al. [2007] analyse the computational complexity of many decision problems occurring within QCGs.

Since their introduction QCGs have been extended in a number of ways. For example, QCGs assume that conflicts of interest do not arise between agents, Dowell et al. [2007] investigate the implications of removing this assumption. In their formulation, Dowell et al. [2007] find that under these circumstances QCGs no longer exhibit the property of envy-freeness so, should utility be transferable, agents may make bribes.

Ågotnes et al. [2006a] present a formal logic for expressing the properties of a QCG and also extend QCGs into a temporal form where games are iterated over time, the former providing the basis for the latter. In their characterisation of temporal qualitative coalitional games (TQCGs) Ågotnes et al. consider that both the goal sets of the agents, and the feasible choices available to agents may vary over time. Although a handful of possibilities for temporally-stable solution concepts are outlined these are not treated in detail; the authors suggest that further study is required in this area [Ågotnes et al., 2009].

Dunne and Wooldridge [2004] consider that agents may have preferences over the goals contained within their individual goal set. Solution concepts for QCGs with preferences are concerned with strong and weak notions of coalitional preference. A coalition is said to strongly prefer some set of goals where that set contains at least one goal, common to all members of the coalition, that is more preferable for each member than any other feasible goal. In the weaker form the requirement for the most preferable goal to be common to all member-agents is relaxed [Dunne and Wooldridge, 2004]. QCGs with preferences are also considered in a temporal form in [Cheng et al., 2010].

2.3 Preference Aggregation in Multi-agent Systems

The question of how to 'fairly' aggregate heterogeneous individual preferences into an overall social preference is central to the problem considered by social choice theory [Arrow, 1950]. Research in this field considers preference aggregation through the study of *social choice functions.* Such functions expect as their input a set of preference relations, representing the preferences of each individual with respect to some finite set of choices, or candidates. The return is a single (social) preference relation derived from those supplied and in accordance with certain desirable qualities.

Arrow [1950] gives five criteria which any fair social choice function should meet. The first of these, known as *unrestricted domain* states that any social choice function should be defined for any combination of individual preferences; the second, *positive association*, requires the monotonic response that, where an individual strengthens their preference for some alternative, that alternative should not be any less preferred in the consequent social preference. The condition of *independence of irrelevant alternatives* requires that if some choice, a, is preferred to some choice, b, the introduction of an alternative choice, c, should not affect this pre-existing preference. A social choice function meets the requirement of *non-imposition* provided that there is no social preference which cannot be reached given an appropriate set of individual preferences, whilst the requirement of *non-dictatorship* stipulates that aggregate social preferences should not solely reflect the preferences of any single individual. Arrow [1950] proceeds to show the impossibility of simultaneously satisfying all of these criteria for all combinations of individual preferences where the number of available choices is greater than two.

Developed independently, the Gibbard-Satterthwaite Theorem [Gibbard, 1973; Satterthwaite, 1975] shows that where the number of choices exceeds two no social choice function exists that can be guaranteed to simultaneously satisfy the constraints of being non-dictatorial and non-impositional without being susceptible to strategic manipulation. That is, by untruthfully reporting their preference an individual can affect the social preference in their favour.

The practical implications of the Gibbard-Satterthwaite result are very much dependent on the circumstances in which the decision is to be made. For an individual to conduct a successful manipulation it is necessary that, in addition to their own preferences, the individual has complete and accurate knowledge of both the algorithm of the social choice function and the preferences of all other participants [Dowding and van Hees, 1977]. In many practical voting situations it may be infeasible for any individual to attain the level of information required in order to effect successful manipulation [Nurmi, 2002]. However, were this information to be available then Conitzer and Sandholm [2006] show that under reasonable assumptions it is impossible to design a social choice function that is usually (frequently often) hard to manipulate.

2.4 Collective Action in Multi-agent Systems

If agents are to be able to act collectively in pursuit of some common goal it is necessary that agents are able to reason not only about their own actions, but also about the actions of other agents [Grosz and Kraus, 1999]. For example agents must coordinate their individual activities such that they do not place resources into contention (the resource allocation problem), repeat the work of others, or omit some crucial objective (the task allocation problem).

These problems are considered in a formalisation known as SharedPlans [Grosz and Kraus, 1996, 1999; Grosz and Sidner, 1990]. In SharedPlans the focus is on creating logical specifications for the mental state of agents when required to align their individual plans with those of other agents attempting to realise a shared goal, typically under constraints of time. In SharedPlans there is an assumption that agents are able to reach an agreement, however in their original formulation no such mechanism is explicitly defined, this is addressed in [Hunsberger and Zancanaro, 2000].

Contemporaneous to the work of Grosz and Sidner, and adopting a similar approach to the problem of collective action Cohen and Levesque [1991]; Levesque et al. [1990] propose the notion of joint intentions, which, like SharedPlans considers that agents acting together do so in the context of a shared, mutual mental state. However, the stance that collective intention can be reduced to a collection of individual intentions has been challenged. Most notably, Searle [1990] argues that no degree of individual intention or commitment can capture the cooperative nature of collective action. Searle compares a number of individuals simultaneously converging at a shelter in order to stay out of the rain, with the same action performed by a group of choreographed dancers. In the first instance, Searle claims that even where each individual holds beliefs about its own actions and the actions of others there is no sense of collective action, whereas, in the latter case the intentions of each individual are in some-way derivable from a collective, or “we intention” a term appropriated from philosophical literature [Tuomela and Miller, 1988].

Tambe [1997] borrows from both the SharedPlans of Grosz and Sidner and the joint intentions of Levesque et al. to create STEAM, a general, and implementable model of agent teamwork. In Tambe’s approach agents commence cooperative behaviour by enacting a protocol to establish commitments. The protocol requires that there is consensus, all agents agree, to adopt a proposed joint persistent goal. Tambe acknowledges that in stipulating absolute agreement prior to establishing a common objective a single agent may prevent the group from embarking upon collective action, and whilst suggesting that negotiation may be useful in resolving differences of opinion the central issue, of whether it is possible to achieve absolute agreement amongst agents, is left open for future consideration.

The principal contribution of the works detailed in this section has been to examine how autonomous agents may function as a team to collectively achieve some common goal. However, it is not generally the case that agents are required to have common goals in order that they embark upon a collective act. Little attention has been given to the selection of individual actions or goals to be adopted or performed by the group as a whole [Sellers et al., 2007] or the definition of mechanisms by which this may be achieved [Zappala, 2008].

2.5 Threshold Models for Sociological and Biological Systems

Threshold models have been found to describe a variety of social phenomena including segregation in urban housing [Schelling, 1969], and the adoption of consumer trends [Granovetter and Soong, 1986]. Granovetter [1978] describes a model in which individuals are faced with the binary decision of whether or not to participate in a riot. In this model each individual's threshold represents the minimum proportion of others which must participate in a riot in order that the given individual will also participate. A population of n individuals is considered with uniformly distributed thresholds $\{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n-1}{n}\}$; the scenario begins with a single instigator performing some riotous display; this behaviour then incites the next, and so forth until mass rioting ensues. Granovetter [1978] finds that the eventual number of rioters is very sensitive to the presence of individuals with low thresholds.

Chwe [1999] recognised that this sensitivity is dependent on certain assumptions regarding how agents may communicate. Chwe considers threshold behaviours for agents where they may interact locally via a communication graph, and may have knowledge of their neighbours' thresholds. Under these assumptions Chwe [1999] shows that the sensitivity identified by Granovetter [1978] may not hold. This change in behaviour is attributed to the additional knowledge agents may gain about each others' threshold through communication.

Many natural systems, for example animal groups, have been shown to employ threshold behaviours when making group decisions. Such consensus decision making has been explained in terms of quorum sensing and the quorum response [Conradt and List, 2009; Sumpter and Pratt, 2008]. Quorum sensing represents the ability of an individual to sense the proportion of their group that is in some particular state; where this proportion grows to reach some threshold level the quorum response is exhibited. In the quorum response the probability of other individuals also entering this state increases sharply; a positive feedback loop is established, permitting a collective state change to cascade throughout the group [Sumpter and Pratt, 2008].

Empirical evidence from biological studies suggests that quorum sensing is found in many natural systems. Perhaps the most primitive example is found in the bacteria *Vibrio fischeri* which are able to sense the presence of certain molecules, released by conspecifics and diffused within their environment. Once the environmental concentration of the molecule, taken to be proportional to the size of the local population, passes a given threshold the bacterial colony as a whole begin to bio-luminesce [Fuqua et al., 1996].

More complex organisms, such as eusocial insects, have developed more intricate quorum-based mechanisms which utilise this underlying principle. The honey bee, *Apis mellifera*, begins migration to a new nesting location by swarming to a convenient branch and then despatching scout bees to find suitable alternative nest sites. Scouts return to the swarm reporting the quality and distance of each potential new site through a series of gestures and movements referred to as the "waggle dance"; their aim is to recruit other bees to replicate this dance, showing support for a particular site. The greater the number of bees dancing for a particular site the greater the probability that other scouts will also investigate this site [Seeley and Buhrman, 1999]. Quorum sensing takes place at the potential nesting site. Where the number of scouts concurrently present at the site passes a quorum threshold, thought to be in region of ten to fifteen bees, the scouts will return to the swarm and initiate a migration to that site [Seeley and Visscher, 2003].

It is thought that bees employ quorum based decision making as it permits the swarm to arrive at a decision outcome iteratively and so more easily than requiring simultaneous, unanimous consensus. Were the latter to be adopted it would require, at least, that each of the (several hundred) scout bees becomes aware of the qualities of each potential nest site and, in some-way, the scouts collectively reach a unanimous agreement as to which site is most appropriate. Not only would such a process require a considerable degree of effort and coordination amongst the scouts, it would also require significantly more time, something which may negatively affect the well-being of the entire, presently homeless, swarm. [Seeley and Visscher, 2003]

Arguably, however, the more individuals involved in the decision making process the more accurate the decision outcome. Condorcet's jury theorem, which considers a group making the binary choice "guilty or not guilty", predicts that where each individual is more likely to choose correctly than incorrectly, the probability that the group as a whole will choose correctly approaches one as the number of individuals approaches infinity. Given this, it may be interesting to understand if quorum-based mechanisms provide speedy outcomes at the expense of accuracy; a question considered by Franks et al. [2003] who examined quorum decision making in ant colonies.

The ant *Temnothorax albipennis* is known to employ a quorum decision making mechanism similar to that of the honey bee when faced with selecting a new nesting site. Rather than the waggle dance, scout ants recruit support for a new site by leading other ants to that site - a procedure known as 'forward tandem running'. As with bees, quorum sensing occurs at the new site; unlike bees, a quorum is established once the rate at which nest mates are encountered reaches some threshold [Pratt, 2005]. In Franks et al. [2003] experimental colonies are encouraged to seek out a new nesting location under either benign or stressful conditions; stressful conditions include (careful) disruption of the current nest site or the simulation of predators. Franks et al. report that in stressful conditions the ants employ a lower quorum threshold and so choose between alternatives more quickly than found in their benign experiments; where the quality of available alternatives differ, decisions made more swiftly were found to be less accurate. Also studying ants of the genus *Temnothorax* Pratt and Sumpter [2006]; Sumpter and Pratt [2008] develop a model of the quorum response in which the probability of accepting an alternative increases sharply once the quorum threshold is reached. In simulations this model predicts that variation in either the threshold itself, or the probability of acceptance permits tuning of the quorum response to favour either the speed of decision making or the accuracy of the decision outcome.

Quorum-based decision making has also been described in terms of the independence of, and interdependence between decision makers. Intuitively, a highly independent individual acts primarily in accordance with its own information, disregarding the preferences or state of others, whereas a highly interdependent individual acts primarily in accordance with its peers. These conditions are respectively analogous to individuals having a very high, or very low quorum threshold.

List et al. [2009] develop an agent-based simulation modelled on decision making in the honey bee to investigate the relationship between independence and interdependence. Systematic variation, at the population level, of the degree to which the agents are independent, or otherwise, reveals that populations of non-interdependent individuals are incapable of reaching a single decision outcome, whereas populations of highly interdependent individuals reach poor (in terms of accuracy) decision outcomes. Accurate decisions, in this case, selection of the highest quality nesting site from five alternatives, only present in populations where individuals are bestowed with a balance of independent/interdependent behaviour. The model of List et al. predicts that accuracy is maintained across a broad range of non-extremal settings for independence/interdependence, even where individuals possess unreliable information. This demonstrates the beneficial effects of information pooling, aggregating information from all decision makers, present in quorum-based decision making.

Agent-based modelling has also been used to investigate threshold decision making in more complex species; Sellers et al. [2007] introduce a simulation of group decision making in chacma baboons (*Papio hamadryas ursinus*). Individuals are modelled with a bio-realistic physiology and a representation of drives and desires for key activities including eating, drinking, socialising and resting. Agents are situated in a complex and dynamic environment which, amongst many features, exhibits realistic monthly variation in food availability. A voting mechanism, with variable threshold, is employed by the group to decide the direction of group travel. Sellers et al. show that this threshold is the single most significant determinate of 'successful' simulations (those in which all agents achieve realistic budgets for the key activities). Subsequent development of this model has considered the effects of environmental conditions upon threshold-based decision making. Hill et al. [2011] show that in rich environments, where food resource is plentiful, a broad range of values for the quorum threshold, including the sub-majority range where less than fifty percent of individuals must agree, lead to successful outcomes for the group. Conversely, where resource is less easily obtained sub-majority decision making proves less effective. Intuitively, where the probability of selecting a poor outcome is low the decision strategy employed by the group is less critical than where the probability of making the 'wrong' decision is high. The latter requires a greater degree of commitment from the group, and hence a higher quorum threshold.

2.6 Threshold Models in Game Theory

In threshold models an individual makes a particular choice only where at least a threshold number of others have first made that same choice. Fundamentally, threshold behaviours can be thought of as imitative behaviours; doing something because some other is doing it. Such behaviours have been studied from a game theoretic perspective. In imitation games [McLennan and Tourky, 2010] two players take the roles of leader and follower; the follower is motivated to act in consensus with the leader.

Threshold behaviours are also the subject of the El Farol problem in which players must decide whether to stay at home or visit a local night-spot. Players will enjoy being at the El Farol bar only if the number of other players also visiting the bar is less than some threshold (typically two-thirds). Where the threshold is exceeded, the bar becomes too crowded, it would have been better for the players to stay at home.

The El Farol problem was the inspiration for minority games. In minority games (MGs) [Challet and Zhang, 1997] players are faced with a binary choice; a payoff is awarded to those players choosing the minority favoured option. Since the objective of this research is to understand how the historic performance of strategies of a single agent might influence that agent's strategy choice in the current game MGs are typically studied in their repeated form.

The conditioning of behaviour given the behaviour of others has been studied in, e.g., congestion games [Rosenthal, 1973], a type of potential game [Monderer and Shapley, 1996]. In these games players do not seek to reach a consensus, instead they wish to minimise costs whilst utilising resources. The cost to each player for the use of each resource is a function of the number of other players also choosing a strategy requiring that resource. It has been shown [Rosenthal, 1973] that every congestion game possesses at least one Nash equilibrium in pure strategies.

Threshold behaviours are also found in anonymous games [Daskalakis and Papadimitriou, 2007] in which the individual utility of participation in some coalition is independent of the identities of the agents concerned. In such situations other factors, including the size of the coalition become determinants of an agent's choice.

2.7 Threshold Models in Computer Science

The potential for threshold mediated behaviours has not been overlooked by the field of computer science. Krasnogor et al. [2005] show how the bacterial quorum sensing model can function as a NAND gate. Bernardini et al. [2007] present a computational model of quorum sensing (again, bacterially inspired) which proves to be equivalent to a Turing machine.

Applications for threshold mediated behaviours are also apparent. Peysakhov and Regli [2005] show that ant-like quorum sensing behaviour can be used to manage services and resources within mobile ad-hoc networks (MANETs). This includes collective decisions such as exclusion of a (potentially) compromised host from the network. Due to their inherently dynamic and unstable nature complete traversal of MANETs is rarely advisable; instead Peysakhov et al. [2006] use a quorum-based approach to determine when enough hosts have been sampled in order to predict the decision outcome with some degree of confidence. The threshold of 'enough' hosts is established by using the binomial probability distribution to determine the probability that k successes ('yes' votes) would be found in n unbiased trials. Where this probability lies outside some confidence interval this would indicate that the population is biased towards one, or other outcome. They show that such an approach reduces both time and bandwidth requirements in comparison to full network traversal with negligible effects on decision accuracy.

In computer security, worms are often malicious, self propagating computer programs; to avoid detection it is desirable that a worm ceases propagation once some critical number of hosts are infected. Vogt et al. [2007] show that quorum sensing is a viable mechanism to create a self-stopping worm which will cease propagation once some quorum threshold is reached. Beckmann et al. [2009] propose a suitable defence strategy against such a worm is to introduce mutants into the population in order to disrupt the process of quorum sensing, an approach known as 'quorum quenching'. Surprisingly, Beckmann et al. find that this approach can be effective where concentration levels of mutants is as low as ten percent; a result which suggests quorum-based decision making procedures may be susceptible to manipulative behaviour.

Perhaps the closest research strand to the subject of this thesis has been inspired by the rapid growth of social networking and viral marketing, especially in the online space. Much of this research has considered the effects of network topology upon individuals which make binary decisions based upon thresholds. Kempe et al. [2003] consider the diffusion of ideas through a network of agents. In their model agents will adopt an idea only if a threshold number of neighbouring agents are also adopters. Their interest is in identifying agents which, based on network location, are most influential. Although this problem is $NP\text{-}hard$ they provide algorithms for approximate solutions. Also interested in threshold mediated propagation via networks, Watts [2002] models random networks of agents with random thresholds. Interest in this is motivated by a desire to understand the effects of heterogeneity upon adoption rates. Watts finds that the likelihood of widespread adoption increases with more heterogeneous thresholds, but decreases with a more heterogeneous network structure.

2.8 Model Analysis Tools

An objective for this research is to create models that are open to study using established methods. One, well established, method for analysing models is the use of model checking tools [Clarke et al., 1994]. Such tools are used to construct a representation of the model as a finite state transition system; the tool can then be used to verify properties of interest exhibited by the model.

The utility of model checking tools to verify properties of complex distributed systems is aptly demonstrated by the work of Havelund et al. [2001] who investigated properties of the Executive Support Language (ESL). ESL is a multi-threaded programming language created by NASA and used in the construction of autonomous machines such as robots and space craft. Using the SPIN model checker [Holzmann, 1997], Havelund et al. identified several errors in ESL including one significant design flaw.

The model checking problem is challenging since models of even a modest practical system may require an enormous amount of states. This is known as the *state explosion problem*. A naive approach to model checking would utilise an *explicit* representation of the state space. Verification of a property, for example: “will the model encounter a deadlock”, requires a (potentially exhaustive) search of the transition graph. Consequently, naive approaches to model checking are rarely found.

In contrast to the explicit representation many model checking tools employ a *symbolic* representation of the state space. Burch et al. [1992] propose a method for succinct (in many but not all cases) representation of finite state transition systems utilising binary decision diagrams [Bryant, 1986]. Burch et al. report that, when using symbolic representations, properties of models with in excess of 10^{20} states can be verified.

In a typical use case for model checking tools the model is first encoded using a model specification language; the choice of language is tool-dependent. Supported languages include general programming languages such as Java [Corbett et al., 2000; Havelund and Pressburger, 2000] as well as languages specific to the model checking domain; these include: Interpreted Systems Programming Language (ISPL) [Lomuscio et al., 2009], Reactive Modules [Alur and Henzinger, 1999] and the Process Meta Language (PROMELA) [Holzmann, 1997].

Having constructed the model, properties of interest are then encoded using a property specification language; again the choice of language is tool-dependent. Typically, however, properties are specified as formulas using a modal temporal logic such as linear temporal logic (LTL) [Pnueli, 1977] or computation tree logic (CTL) [Clarke et al., 1986]. Such logics are used to enquire about properties of the model such as liveness (eventually, will some property hold?) or path invariance (does some property always hold?).

For many model checking applications the dichotomous nature of logic is sufficient to specify properties of interest. However, given the generally non-deterministic behaviour of transition systems this *qualitative* approach can be limiting. In the applied portion of this work (Chapter 7) we use a model checking tool to verify properties of an instance of our temporal model. Whilst we are interested in qualitative properties, such as: “can the agents maintain consensus?” we are also interested *quantitative* properties, such as: “how likely is it that the agents will maintain consensus?”.

Probabilistic real-time computation tree logic (PCTL) Hansson and Jonsson [1994] extends CTL to include probabilistic operators. The logic is particularly suited to analysing processes that can be modelled as a Markov decision process (MDP). Formula in PCTL permit property specifications of the form: “what is the probability that some property will hold?”. PCTL is supported by only a subset of model checking tools, these include the Approximate Probabilistic Model Checker (APMC) [Herault et al., 2006], the Markov Reward Model Checker (MRMC) [Katoen et al., 2011] and PRISM [Kwiatkowska et al., 2011].

For our applied work we chose to use the PRISM model checking tool. This choice was largely determined by our technical need to verify properties specified in both probabilistic and non-probabilistic logics. We found PRISM to be well suited to this requirement as it supports property specification in a variety of logics including CTL, LTL and PCTL. Several pragmatic concerns also aided our choice; these include the existence of an active development/support community and that we found the tool to be intuitive and hence present a shallow learning curve.

Chapter 3

Consensus Games

3.1 Introduction

In this chapter we present a novel model of consensus decision making for multi-agent systems. Formulated as a game, our model, that of Consensus Games (CGs), is inspired by both the quorum mechanisms found in biological systems and the threshold models developed in sociological literature. A CG describes a group of agents seeking consensus regarding the formation of coalitions.

Consensus games extend the threshold model proposed by Granovetter [1978] beyond binary choice decisions to the more general problem of coalition formation. For each coalition of which an agent may be a member, each agent holds a threshold representing the proportion of agents from that coalition which must support the formation of the coalition in order that the agent will also support the coalition. There is consensus about the formation of a particular coalition only where all member-agents support the formation of that coalition.

3.2 Consensus Games

We begin by formally defining a consensus game.

Definition 3.1. *A consensus game (CG) is a tuple* $\Gamma = \langle A, q \rangle$ *where:*

A *is a finite set of agents,* $\{1, \ldots, n\}, n \geq 2$.

q *is a quorum function. It is a partial function* $q : A \times 2^A \rightarrow [0, 1]$ *which takes an agent* $i \in A$ *and a coalition of agents* C *such that* $i \in C$ *and returns a number in the interval* $[0, 1]$.

The value of the quorum function for a coalition indicates the agent's 'degree of support' for the formation of that coalition. For an agent $i \in C \subseteq A$ the quorum function $q(i, C)$ gives the minimum proportion of agents in C which must support the formation of the coalition C in order that i will support the formation of this coalition. Where $q(i, C) = 0$ agent i unconditionally supports the formation of the coalition C, where $0 < q(i, C) \leq \frac{|C|-1}{|C|}$ the agent conditionally supports formation of the coalition C; where $\frac{|C|-1}{|C|} < q(i, C) \leq 1$ the agent objects to the formation of the coalition C.

We use the abbreviation $q\#(i, C)$ to denote the number of other agents from C which have to support C in order for i to support C. Formally, $q\#(i, C)$ is the minimal natural number k such that $q(i, C) \leq k/|C|$. Lastly, $n_k(C)$ denotes the number of agents $i \in C$ with $q\#(i, C) = k$.

3.3 Strong Consensus

Now we define the main solution concept of consensus games, a strong consensus coalition. A *strong consensus coalition* $C \subseteq A$ is a coalition where for each agent $i \in C$ the quorum threshold $q(i, C)$ is satisfied in the sense that C contains at least $q\#(i, C)$ other agents with strictly lower $q\#$ values.

Definition 3.2. *A coalition $C \subseteq A$ is a strong consensus coalition if the following conditions hold:*

- $n_0(C) \neq 0$
- *if* $n_k(C) \neq 0$*, then* $\sum_{j<k} n_j(C) \geq k$

In a strong consensus coalition there exists at least one $i \in C$ such that $q(i, C) = 0$ and all agents' thresholds are satisfied. Note, the definition implies that if C is a strong consensus coalition, then $n_{|C|}(C) = 0$. In a strong consensus coalition no agent objects to formation of the coalition.

Consider the following example.

Example 3.1. *Alice (Al) and Bob (Bo) are considering whether to get married. Bob no longer wishes to be a bachelor and is keen to be married. Alice is not opposed to the idea of marrying Bob provided that she knows that Bob also wants to marry her, otherwise Alice will happily continue her single lifestyle. Alice's and Bob's positions can be formalised as the consensus game* $\Gamma_1 = \langle A, q \rangle$ *where:*

$A = \{Al, Bo\}$

$$q(i,C) = \begin{cases} 0 & \textit{if } i = Bo \textit{ and } C = \{Al, Bo\} \\ 0.5 & \textit{if } i = Al \textit{ and } C = \{Al, Bo\} \\ 1 & \textit{if } i = Bo \textit{ and } C = \{Bo\} \\ 0 & \textit{if } i = Al \textit{ and } C = \{Al\} \end{cases}$$

In Example 3.1, Bob unconditionally supports the formation of the grand coalition (of all agents); Alice conditionally supports formation of this coalition provided that one other agent (Bob) also supports forming this coalition. Alice also unconditionally supports formation of the singleton coalition $\{Al\}$, whereas Bob objects to the formation of the singleton coalition $\{Bo\}$. The grand coalition in this example is a strong consensus coalition.

Next, we show that there is an alternative definition of a strong consensus coalition as a fixed point of a function. This definition intuitively corresponds to agents indicating their support for a coalition.

Consider the following function $f_C : 2^A \to 2^A$ defined relative to $C \subseteq A$:

$$i \in f_C(Q) \text{ iff } i \in C \text{ and } |Q \cap C \setminus \{i\}| \geq q(i,C) \times |C|$$

This function takes as its input a set $Q \subseteq A$ and returns the set of agents in C whose quorum thresholds are satisfied in the sense that for each agent $i \in f_C(Q)$, $q_{\#}(i,C) \leq |Q \cap C|$. If $Q = \emptyset$, f_C will contain only the agents i with $q(i,C) = 0$, if Q is the set of agents which have unconditional support for C, then $f_C(Q)$ will contain the agents i with $q\#(i,C) \leq |Q|$, and so forth.

To show that a coalition C is a strong consensus coalition if and only if it is the least fixed point of f_C the following auxiliary result is required:

Proposition 3.1. *The function f_C is guaranteed to possess at least one fixed point.*

Proof. Existence of at least one fixed point is guaranteed for monotonic functions; Knaster-Tarski theorem [Tarski, 1955].

For monotonicity it must be shown that $\forall Q, Q' \subseteq A$ where $Q \subseteq Q'$ it is always the case that $f_C(Q) \subseteq f_C(Q')$.

Let $Q \subseteq Q'$. It must be shown that for every i, if $i \in f_C(Q)$, then $i \in f_C(Q')$. Assume $i \in f_C(Q)$. By the definition of f_C, $i \in C$ and $|Q \cap C \setminus \{i\}| \geq q(i,C) \times |C|$. Since $Q \subseteq Q'$, $|Q' \cap C \setminus \{i\}| \geq |Q \cap C \setminus \{i\}|$ hence $|Q' \cap C \setminus \{i\}| \geq q(i,C) \times |C|$ and $i \in f_C(Q')$. □

The least fixed point of a function can be established by recursive calls to the function starting with the empty set (of agents) as an argument; each invocation of f_C will be referred to as a *round*. If C can achieve strong consensus, then it will be achieved in at most $|C|$ rounds.

We can now prove:

Theorem 3.1. *C is a strong consensus coalition if and only if it is the least fixed point of f_C.*

Proof. Assume C is a strong consensus coalition. Then $n_0(C) \neq 0$ and in the first round $|f_C(\emptyset)| = |\{i \in C : q\#(i, C) = 0\}| = n_0(C)$, so $f_C(\emptyset) \neq \emptyset$.

Denote the k^{th} application of f_C as f_C^k; at each round $k > 1$, $f_C^k(\emptyset) = f_C^{k-1}(\emptyset) \cup \{i \in C : q\#(i, C) \leq |f_C^{k-1}(\emptyset)|\}$

By the definition of a strong consensus coalition, additional agents support at each round until all agents support. The set $\{i \in C : q\#(i, C) \leq |f_C^{k-1}(\emptyset)|\}$ is always non-empty until $f_C^{k-1}(\emptyset) = |C|$, so C is the least fixed point of f_C.

Assume C is the least fixed point of f_C. Then the first condition in the definition of a strong consensus coalition is satisfied because $f_C(\emptyset) \neq \emptyset$ hence there are agents i with $q(i, C) = 0$. To show that for every k, if $n_k(C) \neq 0$, then there are at least k agents in C with strictly lower $q\#$ values, consider an agent i with $q\#(i, C) = k$. Since C is the least fixed point of f_C, at some round, m, $i \in f_C^m(\emptyset)$. This means that $|f_C^m(\emptyset)| \geq k$, and since $|f_C^m(\emptyset)|$ is the number of agents with lower $q\#$ values, $\sum_{j<k} n_j(C) \geq k$. □

3.4 The q-Minimal Core

The core is a popular solution concept in game theory, aggregating equilibria that are both individually and collectively rational. In traditional, quantitative game theoretic models, rational behaviour is typified by agents maximising some notion of utility. Agents are said to act with individual rationality where each agent will achieve their maximum reward given the behaviour of the other agents; this situation corresponds to the Nash equilibrium of the game. Collectively rational outcomes are those where no subset of agents can achieve a higher reward through defection.

By contrast, CGs are qualitative; they do not describe a pay-off or reward structure. Instead, individually rational behaviour is associated with the quorum thresholds of the agents for the coalitions of which they may be a member. It is individually rational for an agent to support the formation of some coalition only if the number of agents already known to be supporting that coalition is a least as great as the proportion specified by the agent's quorum threshold for that coalition.

Notions of equilibria constituting collectively rational outcomes for CGs are harder to define. One natural way of distinguishing between coalitions is in terms of the effort required to reach consensus. For some coalitions, strong consensus may be established in a single round, while for others it may require as many as $|C|$ rounds. The number of rounds required to reach consensus for a given coalition can be taken as a measure of the ease with which the agents reach consensus and hence the stability of the resulting coalition. Let $rounds(C)$ be the number of rounds necessary for strong consensus to be established for some coalition $C \subseteq A$.

Definition 3.3. *The coalition $C \subseteq A$ is a q-minimal consensus coalition only if:*

- *C is a strong consensus coalition,*
- *there exists no $C' \subset C$ such that C' is also a strong consensus coalition and $rounds(C') < rounds(C)$*

The q-minimal core aggregates q-minimal consensus coalitions.

3.5 Complexity of CGs

Our characterisation of the computational complexity of CGs considers three natural decision problems. These decision problems address fundamental questions for CGs regarding the presence or otherwise of equilibria within a game. The first considers the problem of verification, the second, that of existence whilst the third considers the matter of non-existence:

Consensus Coalition (CC): Can a given coalition reach consensus?

Consensus Coalition Exists (CE): Does there exist some coalition which can reach consensus?

No Consensus Coalition (NC): Is there no coalition which can reach consensus?

These decision problems are considered, first for strong consensus coalitions and subsequently for q-minimal consensus coalitions. We begin by establishing the representational scheme and abstract computational machine for these analyses.

3.5.1 Representation

When considering the representation of CGs the structure of most interest is the quorum function, q. The quorum threshold has to be specified for each coalition $C \in \mathcal{P}(A) \setminus \emptyset$ and for each agent $i \in C$. To do this, each coalition $C \subseteq A$ is represented as a set of pairs $(i, q(i, C))$. The overall representation is the set $R = \{rep(C) \mid C \in \mathcal{P}(A) \setminus \emptyset\}$, where $rep(C) = \{(i, q(i, C)) \mid i \in C, q(i, C) \in [0, 1]\}$. Note that the size of R is exponential in the number of agents, n.

We assume that R is implemented as a random access data structure, hence, the following results are given for the non-deterministic random access machine (NRAM) model of computation [van Emde Boas, 1990].

3.5.2 Complexity of strong consensus decision problems

STRONG CONSENSUS COALITION (SCC)

Given a CG $\Gamma = \langle A, q \rangle$ and a coalition $C \subseteq A$, can C reach strong consensus?

A deterministic algorithm must verify that C is the least fixed point of f_C. Algorithm 3.1 runs in time which is polynomial (linear) in n and therefore lies within $P(n)$. Note that we can also obtain an $O(n \times log(n))$ algorithm, which runs in constant space, by sorting C on $q(i, C)$.

Algorithm 3.1 Can C reach strong consensus.

```
function SCC(R, C)
    array support[|C| + 1] ← {0, ..., 0}
    for all (i, q) ∈ C do
        k ← ⌈q × |C|⌉
        support[k] ← support[k] + 1
    end for
    s ← support[0]
    for k from 1 to |C| do
        if k ≤ s then
            s ← s + support[k]
        else
            return false
        end if
    end for
    return true
end function
```

STRONG CONSENSUS COALITION EXISTS (SCE)

Given a CG $\Gamma = \langle A, q \rangle$ is there some $C \subseteq A$ which can reach strong consensus?

A deterministic algorithm would iterate over R, checking for each $C \subseteq A$ whether C is a strong consensus coalition. Hence the problem is in $O(n2^n)$.

A non-deterministic algorithm first guesses an index of a coalition $C \in R$ and then checks that C can reach strong consensus. This can be done in time linear in n using Algorithm 3.1. This gives a non-deterministic linear time algorithm for a random access machine. Hence, the problem is in $NP(n)$ for NRAM.

NO STRONG CONSENSUS COALITION (SNC)

Given a CG $\Gamma = \langle A, q \rangle$ is there no $C \subseteq A$ which can reach strong consensus?

A deterministic algorithm must verify that $\neg\exists C \subseteq A$ such that C is the least fixed point of f_C. Hence the problem is in $O(n2^n)$.

The problem of verifying that there exists some coalition which can reach strong consensus is in $NP(n)$. Therefore the complement of that problem, verifying that there exists no coalition which can reach strong consensus is in $co\text{-}NP(n)$ for NRAM.

q-MINIMAL STRONG CONSENSUS COALITION (QM-SCC)

A deterministic algorithm must verify that C is the least fixed point of f_C and that $\neg\exists C' \subset C$ such that C' is the least fixed point of $f_{C'}$ and C' reach consensus in strictly fewer rounds than C. Algorithm 3.2 computes the number of rounds required to encounter the least fixed point of f_C. The algorithm has time complexity $O(n)$ and so is in $P(n)$.

Algorithm 3.2 Number of rounds for C.

```
function rounds(rep(C))
    array support[|C| + 1] ← {0, . . . , 0}
    for all (i, q) ∈ C do
        k ← ⌈q × |C|⌉
        support[k] ← support[k] + 1
    end for
    r ← 0
    i1 ← 1
    i2 ← s ← support[0]
    while i1 ≤ i2 do
        r ← r + 1
        for k from i1 to i2 do
            s ← s + support[k]
        end for
        i1 ← i2 + 1
        i2 ← s
    end while
    return r
end function
```

Algorithm 3.3 then verifies that a given coalition, C is a q-minimal strong consensus coalition, by iterating over all subsets of C. Hence the problem is in $O(n2^n)$.

A non-deterministic algorithm to solve the complement of this problem (decide whether a coalition is *not* a q-minimal consensus coalition) first checks whether C is a strong consensus coalition (and returns true if it is not); if C is a strong consensus coalition, it will guess an index of an coalition $C' \subset C$ and check that C' is a strong consensus coalition which converges in fewer rounds than for C (and returns false if it does). So the problem of deciding whether a coalition is *not* a q-minimal strong consensus coalition is in NP(n) on NRAM. Hence deciding whether a coalition is a q-minimal strong consensus coalition is in $co\text{-}NP(n)$ for NRAM.

Algorithm 3.3 Is C a q-minimal strong consensus coalition.

```
function QM-SCC(C, R)
    if ¬SCC(R, C) then
        return false
    end if
    for all C′ ⊂ C ∈ R do
        if SCC(R, C) ∧ rounds(C′) < rounds(C) then
            return false
        end if
    end for
    return true
end function
```

q-MINIMAL STRONG CONSENSUS COALITION EXISTS (QM-SCE)

A deterministic algorithm must verify that $\exists C \subseteq A$ such that C is the least fixed point of f_C and that $\neg\exists C' \subset C$ such that C' is the least fixed point of $f_{C'}$ and C' reach consensus in strictly fewer rounds than C. Hence the problem is in $O(n2^n)$.

If there exists some coalition which is a strong consensus coalition then either that coalition itself, or some subset of that coalition, will be a q-minimal strong consensus coalition. Therefore, this decision problem can be solved in a manner similar to SCE and so is in $NP(n)$ for NRAM.

NO q-MINIMAL STRONG CONSENSUS COALITION (QM-SNC)

A deterministic algorithm must verify that $\neg\exists C \subseteq A$ such that C is the least fixed point of f_C and that $\neg\exists C' \subset C$ such that C' is the least fixed point of $f_{C'}$ and C' reach consensus in strictly fewer rounds than C. Hence the problem is in $O(n2^n)$.

If there exists some coalition which is a strong consensus coalition then either that coalition itself, or some subset of that coalition will be a q-minimal strong consensus coalition. Therefore, this decision problem can be solved in a manner similar to SNC and so is in *co*-$NP(n)$ for NRAM.

3.6 Weak Consensus

The results from the previous section suggest that strong consensus is reasonably easy (takes time linear in the number of agents) to reach. However, it may seem that there is a notion of consensus that is even easier to reach, and which has the same intuitive appeal as strong consensus. Weak consensus is a notion of consensus corresponding to no agent *objecting* to the formation of a coalition.

Definition 3.4. *C is a weak consensus coalition if no agent $i \in C$ has $q(i, C) > \frac{|C|-1}{|C|}$.*

Interestingly, this notion of consensus also has a fixed point characterisation:

Theorem 3.2. *C is a weak consensus coalition iff C is the greatest fixed point of f_C.*

Proof. The only if direction: if there exists $i \in C$ such that $q(i, C) > \frac{|C|-1}{|C|}$, then by the definition of f_C, $i \notin f_C(C)$ so $f_C(C) \neq C$.

For the if direction, assume that for all $i \in C$, $q(i, C) < \frac{|C|-1}{|C|}$. Observe that $f_C(C) = C$; by the definition of f_C, any $i \in f_C(C)$ iff $i \in C$ and $q(i, C) \leq \frac{|C|-1}{|C|}$. Since the latter holds for all $i \in C$: $i \in f_C(C)$ iff $i \in C$ hence C is a fixed point.

It is also the greatest fixed point, since the definition implies that $f_C(C') \subseteq C$ for any C', so $f_C(C') \subset C'$ for any $C' \supset C$. □

An example is called for:

Example 3.2. *Consider again Example 3.1. However, now assume that both Alice and Bob conditionally support getting married. Alice's and Bob's positions can be formalised as the consensus game $\Gamma_2 = \langle A, q \rangle$ where:*

$$A = \{Al, Bo\}$$

$$q(i, G') = \begin{cases} 0.5 & \textit{if } i = Al \textit{ and } G' = \{Al, Bo\} \\ 0.5 & \textit{if } i = Bo \textit{ and } G' = \{Al, Bo\} \\ 1 & \textit{otherwise} \end{cases}$$

In Example 3.2, both Alice and Bob will support formation of the grand coalition, and so get married, if the other also supports this. As neither Alice nor Bob object to the formation of the grand coalition, it is a weak consensus coalition.

Finally, it is easy to show that while every strong consensus coalition is also a weak consensus coalition (since no strong consensus coalition can contain an agent that objects), the converse is not the case: there exists weak consensus coalitions which are not strong consensus coalitions. Since $f_C(C') = C$ for any $C' \supseteq C$, it follows that if C is the least fixed point of f_C, then it is also the greatest fixed point of f_C. In Example 3.2 the grand coalition is a weak consensus coalition but is not a strong consensus coalition.

3.6.1 The C-minimal core

The agents in a weak consensus coalition reach consensus in a single round. The analogue of a q-minimal consensus coalition is therefore not informative for weak consensus. Instead we adopt an approach similar to the qualitative model of the core introduced in Wooldridge and Dunne [2004] for qualitative coalitional games (QCG). A coalition is in the qualitative core of a QCG if and only if that coalition is stable and no subset of that coalition is also stable. Consider two stable coalitions, $C \subseteq A$ and $C' \subset C$. Qualitatively speaking, agents in the coalition C' can do no better by forming C' than they would by participating in the larger coalition C; however, by the same rationale agents in C' will do no worse by forming this smaller coalition. In Wooldridge and Dunne [2004] it is argued that the existence of the stable coalition C' undermines the stability of C; there is nothing impelling C to remain together.

Building on this qualitative definition of the core where collective rationality is associated with minimality we have:

Definition 3.5. *The coalition $C \subseteq A$ is a C-minimal consensus coalition only if:*

- *C is a weak consensus coalition,*
- *there exists no $C' \subset C$, such that C' is also a weak consensus coalition.*

The C-minimal core of a CG contains only C-minimal consensus coalitions. The C-minimal core aggregates consensus coalitions which are collectively rational in the sense that they are immune to defection by some agents $C' \subset C$.

3.6.2 Complexity of weak consensus decision problems

In the remainder of this section decision problems for weak consensus and C-minimal consensus coalitions are considered.

WEAK CONSENSUS COALITION (WCC)

This first decision problem considers the complexity of determining if a given coalition is a weak consensus coalition. Given a CG $\Gamma = \langle A, q \rangle$ and a coalition $C \subseteq A$, will C reach weak consensus?

A deterministic algorithm must verify that $\neg\exists i \in C$ such that $q(i, C) > \frac{|C|-1}{|C|}$ (from Theorem 3.2). It simply iterates through $rep(C)$ checking that no agent i has $q(i, C) > \frac{|C|-1}{|C|}$. Hence the problem is in $O(n)$.

WEAK CONSENSUS COALITION EXISTS (WCE)

Given a CG $\Gamma = \langle A, q \rangle$ will any $C \subseteq A$ reach weak consensus?

A deterministic algorithm must verify that $\exists C \subseteq A$ such that $\neg\exists i \in C$ such that $q(i, C) > \frac{|C|-1}{|C|}$. Hence the problem is in $O(n2^n)$.

A non-deterministic algorithm first guesses an index of a coalition $C \in R$ and then checks that C can reach weak consensus. This can be done in time linear in n. This gives a non-deterministic linear time algorithm for a random access machine. Hence, the problem is in $NP(n)$ for NRAM.

NO WEAK CONSENSUS COALITION (WNC)

Given a CG $\Gamma = \langle A, q \rangle$ can no $C \subseteq A$ reach weak consensus?

A deterministic algorithm must verify that $\neg\exists C \subseteq A$ such that $\neg\exists i \in C$ such that $q(i, C) > \frac{|C|-1}{|C|}$. Hence the problem is in $O(n2^n)$.

The problem of verifying that there exists some coalition which can reach weak consensus is in $NP(n)$ (from WCE). Therefore the complement of that problem, verifying that there exists no coalition which can reach weak consensus is in *co*-$NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS COALITION (CM-WCC)

Given a CG $\Gamma = \langle A, q \rangle$ and a coalition $C \subseteq A$, is C a C-minimal weak consensus coalition?

A deterministic algorithm must verify that C is a weak consensus coalition and that $\neg\exists C' \subset C$ such that C' is also a weak consensus coalition. Hence the problem is in $O(n2^n)$.

A non-deterministic algorithm to solve the complement of this problem (decide whether a coalition is *not* a C-minimal weak consensus coalition) first checks whether C is a weak consensus coalition (and returns true if it is not); if C is a weak consensus coalition, it will guess an index of an coalition $C' \in R \setminus C$ and checks that C' is a weak consensus coalition. So the problem of deciding whether a coalition is *not* a C-minimal weak consensus coalition is in NP(n) on NRAM. Hence deciding whether a coalition is a C-minimal weak consensus coalition is in *co*-$NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS COALITION EXISTS (CM-WCE)

Given a CG $\Gamma = \langle A, q \rangle$ is some $C \subseteq A$ a C-minimal weak consensus coalition?

A deterministic algorithm must verify that $\exists C \subseteq A$ such that C is a weak consensus coalition and that $\neg\exists C' \subset C$ such that C' is also a weak consensus coalition. Hence the problem is in $O(n2^n)$.

If there exists some coalition which is a weak consensus coalition then either that coalition itself, or some subset of that coalition will be a C-minimal weak consensus coalition. Therefore, this decision problem can be solved in a manner similar to WCE and so is in $NP(n)$ for NRAM.

NO C-MINIMAL WEAK CONSENSUS COALITION (CM-WNC)

Given a CG $\Gamma = \langle A, q \rangle$ is there no $C \subseteq A$ a C-minimal weak consensus coalition?

A deterministic algorithm must verify that $\neg\exists C \subseteq A$ such that C a C-minimal weak consensus coalition. Hence the problem is in $O(n2^n)$.

If there exists some coalition which is a weak consensus coalition then either that coalition itself, or some subset of that coalition will be a C-minimal weak consensus coalition. Therefore, this decision problem can be solved in a manner similar to WNC and so is in co-$NP(n)$ for NRAM.

3.7 Incomplete Information in CGs

The quorum function, q, encodes the entire information of a CG and is analogous to functions found in characteristic form representations for game theoretic models. In those models it is usual to distinguish between situations where the characteristic function is common knowledge, and those where it is not. Games where the characteristic function is common knowledge (all agents know the characteristic function, and all agents know that all agents know the characteristic function, and ... *ad infinitum*) are referred to as games of complete information, whereas games lacking in such common knowledge are games of incomplete information. This section considers CGs under incomplete information. Our opening assumptions are that agents know only their own quorum threshold(s) and can only communicate their own support, or lack thereof, for the formation of a given coalition.

Consider $A = \{1 \ldots n\}$ and some $C \subseteq A$. For each agent $i \in C$ let s^i_C be a Boolean variable with the intended meaning that i supports the formation of coalition C. Consensus about C is achieved if the agents in C achieve common knowledge that $\bigwedge_{i \in C} s^i_C$ holds. We are going to propose knowledge-based protocols for achieving strong and weak consensus under incomplete information.

Note that as shown by Fagin et al. [1995], no fact can become common knowledge and hence no agreement is possible in practical distributed systems, which have temporal imprecision etc. We make the usual idealising assumptions that: all agents' announcements are delivered reliably (broadcast) and arrive at a tick of a common clock or at a numbered 'round' which is common knowledge among the agents.

3.7.1 Strong consensus

In order to reach strong consensus under incomplete information, it is sufficient that each agent has knowledge of its own quorum threshold. This can be expressed, for each agent, as a formula, $s_{q\#(i,C)}$, which holds where $q\#(i,C)$ other agents from C support C. The formula $s_{q\#(i,C)}$ is a disjunction of statements of the form $s_C^{i_1} \wedge \ldots \wedge s_C^{i_k}$ where $q\#(i,C) = k$, for all possible groups of agents of size k from C which do not include i. If an agent objects, e.g., $q\#(i,C) = |C|$, this disjunction is empty so $s_{q\#(i,C)} = \bot$. If $q\#(i,C) = 0$ (unconditional support), then $s_{q\#(i,C)}$ corresponds to a disjunction of empty conjunctions which is $\top$.

For example, if $C = \{1,2,3\}$ and $q\#(1,C) = 2$, then

$$s_{q\#(1,C)} = s_C^2 \wedge s_C^3$$

If $q\#(2,C) = 1$, then

$$s_{q\#(2,C)} = s_C^1 \vee s_C^3$$

Finally, if $q\#(3,C) = 0$, then $s_{q\#(3,C)} = \top$.

We refer to the protocol for reaching strong consensus as $LFP(C)$. We use the epistemic operator $K_i\Phi$ to denote that agent i knows that Φ. The $LFP(C)$ protocol is very simple: at each round, if $K_i s_C^i$ holds in this round, agent i announces s_C^i. We assume that the agent 'remembers' if it made the announcement at the previous round, and does not repeat it; this minimises the number of messages. When consensus is reached or it is common knowledge that the round is $|C| + 1$, the protocol terminates. The $LFP(C)$ protocol requires that agents can correctly derive s_C^i; therefore the protocol assumes that each agent has the following original knowledge:

- $K_i(s_{q\#(i,C)} \leftrightarrow s_C^i)$

Theorem 3.3. *$LFP(C)$ for C always achieves strong consensus if C is a strong consensus consensus coalition, and requires at most $|C|$ messages.*

Proof. Let the rounds be numbered $1, \ldots, |C|$, and denote the number of agents i announcing s^i_C in round m by $\#m$. In round 1, the agents i which announce s^i_C are exactly those who have $q\#(i, C) = 0$, in other words those in $f_C(\emptyset)$. Consider round m. Clearly, after s^i_C is announced, s^i_C is common knowledge. If there are agents j which require $\#m$ or fewer other agents to support C in order to support it themselves, then in the next round they can derive from $K_j s^i_C$ (for all i which announced s^i_C) and from their original knowledge $K_j(s_{q\#(j,C)} \leftrightarrow s^j_C)$ that $K_j s^j_C$, those agents j will derive s^j_C. Those agents j are in $f^m_C(\emptyset)$. If C is a strong consensus coalition, each agent $i \in C$ will announce s^i_C at some round, and after the last agent does this it is common knowledge that $\bigwedge_{i \in C} s^i_C$. In general, each agent makes at most one announcement, hence the protocol requires at most $|C|$ messages. □

Note that if strong consensus on C is reached, the number of the round when this happens is also common knowledge among C. This gives the agents a measure of the quality of consensus or stability of the coalition as the number of rounds.

The communication complexity (in terms of the number of messages required) of decision problems concerning verification, existence and non-existence of strong consensus coalitions using $LFP(C)$ can now be characterised.

DISTRIBUTED STRONG CONSENSUS COALITION (D-SCC)

Given a CG $\Gamma = \langle A, q \rangle$ and some $C \subseteq A$, can C reach strong consensus?

This decision problem is solved by executing the protocol $LFP(C)$. Each agent in C sends at most one message, so the total number of messages required is $O(n)$.

DISTRIBUTED STRONG CONSENSUS COALITION EXISTS (D-SCE)

Given a CG $\Gamma = \langle A, q \rangle$, can any $C \subseteq A$, reach strong consensus?

This problem is solved by executing the protocol $LFP(C)$ in parallel for all C. The total number of messages is $O(n2^n)$ in the worst case.

Note however, that given the plausible assumption that the number of coalitions each agent supports is 'small', for example, $O(n)$, then the worst case number of messages would be $O(n^2)$.

NO DISTRIBUTED STRONG CONSENSUS COALITION (D-SNC)

Given a CG $\Gamma = \langle A, q \rangle$, can no $C \subseteq A$ reach strong consensus?

The problem can be solved by executing the protocol $LFP(C)$ for each of the $2^n - 1$ coalitions $C \in \mathcal{P}(A) \setminus \emptyset$. The number of messages in the worst case is $O(n2^n)$.

3.7.2 Weak consensus

To reach weak consensus, under incomplete information the agents start with a different representation of their knowledge of their quorum threshold, and follow a different protocol. The $GFP(C)$ protocol for weak consensus is very simple. It consists of a single round, where agents i with $q\#(i, C) = |C|$ announce the proposition o^i_C indicating that they object to the formation of C.

The $GFP(C)$ protocol requires that agents can correctly derive o^i_C and so assumes that the original knowledge of each agent is:

- agents i with $q\#(i, C) = |C|$ know that they object to the formation of C:

$$K_i o^i_C$$

- it is common knowledge that agents in C support the formation of C if and only if none of them objects, we denote by C_C that it is common knowledge for all agents $i \in C$:

$$C_C \bigwedge_{i \in C} (\bigwedge_{j \in C} \neg o^j_C \leftrightarrow s^i_C)$$

- it is common knowledge that if i knows that it objects to the formation of C, then i will announce this in the first (and only) round of the protocol.

Theorem 3.4. *The GFP protocol for C achieves weak consensus iff C is a weak consensus coalition, and requires only one round of announcements (consisting of $|C|$ messages in the worst case).*

Proof. If at least one agent i announces o^i_C, it immediately becomes common knowledge that o^i_C and hence that $\neg \bigwedge_{i \in C} \neg o^i_C$, which using

$$C_C \bigwedge_{i \in C} (\bigwedge_{j \in C} \neg o^j_C \leftrightarrow s^i_C)$$

implies that it is common knowledge that C is not a consensus coalition ($C_C \neg \bigwedge_{i \in C} s^i_C$).

If no agent i announces o^i_C, after the round is over the common knowledge is that $\bigwedge_{i \in C} \neg K_i o^i_C$ and hence $\bigwedge_{i \in C} \neg o^i_C$. From this, the agents derive common knowledge that for every i, s^i_C, hence common knowledge that C is a weak consensus coalition. $\square$

We now consider the complexity of decision problems associated with the $GFP(C)$ protocol.

DISTRIBUTED WEAK CONSENSUS COALITION (D-WCC)

Given a CG $\Gamma = \langle A, q\rangle$ and some $C \subseteq A$, can C reach weak consensus?

This decision problem is solved by executing the protocol $GFP(C)$. Each agent will make at most one announcement, hence this requires $O(n)$ messages.

DISTRIBUTED WEAK CONSENSUS COALITION EXISTS (D-WCE)

Given a CG $\Gamma = \langle A, q\rangle$ can any $C \subseteq A$ reach weak consensus?

The problem can be solved by executing the protocol $GFP(C)$ for each of the $2^n - 1$ coalitions $C \in \mathcal{P}(A) \setminus \emptyset$. The total number of messages is $O(n2^n)$ in the worst case. Note that in the case when each agent supports only few coalitions, this is an unavoidably wasteful way of finding a weak consensus coalition.

NO DISTRIBUTED WEAK CONSENSUS COALITION (D-WNC)

Given a CG $\Gamma = \langle A, q\rangle$ can no $C \subseteq A$ reach weak consensus?

The problem can be solved by executing the protocol $GFP(C)$ for each of the $2^n - 1$ coalitions $C \in \mathcal{P}(A) \setminus \emptyset$. The total number of messages is $O(n2^n)$.

3.8 Revelation of Information in CGs

We have shown that, in CGs, agents can reach either strong or weak consensus without explicit knowledge of each others' thresholds. Consensus can be achieved through public announcements regarding an agent's support or objection (strong consensus and weak consensus, respectively). However, in making such announcements agents may still reveal information regarding their actual threshold value(s). This section considers what an agent may reveal about its threshold value $q(i, C)$, and hence what agents may come to know, about the thresholds of other agents.

3.8.1 Strong consensus

Consider the protocol $LFP(C)$. Let m denote the current round, $1 \leq m \leq |C|$. As before, let $|f_C^m(\emptyset)|$ denote the total number of agents supporting C at the end of round m. Note that $|f_C^{m\leq 0}(\emptyset)| = 0$.

Now, if in some round m agents $i \in C$ announce s_C^i it must be that $|f_C^{m-2}(\emptyset)| < q\#(i, C)$ (otherwise i would have announced s_C^i in the previous round) and $|f_C^{m-1}(\emptyset)| \geq q\#(i, C)$ (otherwise i would not announce s_C^i in the current round).

Proposition 3.2. *In round m agents can construct a bounded inequality about the thresholds of those agents $h \in C$ announcing s_C^h in that round. The $LFP(C)$ protocol creates common knowledge that $\forall h \in f_C^m(\emptyset) \setminus f_C^{m-1}(\emptyset), |f_C^{m-2}(\emptyset)| < q\#(h, C) \leq |f_C^{m-1}(\emptyset)|$.*

Proof. We must show that for agents $h, i \in C, h \neq i$ where in some round $m - 1$, $\neg K_i(s_C^h)$ and subsequently, in round m, $K_i(s_C^h)$, in all states of the world considered possible by agents i at round m, the inequality $|f_C^{m-2}(\emptyset)| < q\#(h, C) \leq |f_C^{m-1}(\emptyset)|$ holds.

Where necessary we will use the notation $K_i^m \Phi$ to mean that in all states of the world considered possible by i after all announcements have been made in round m, Φ holds. Where the round is omitted the intended meaning of $K_i \Phi$ is that Φ holds in all possible worlds in all rounds.

First we define a set of possible worlds, W. Each possible world $w \in W$; represents one possible state of the world at the end of a round of the $LFP(C)$ protocol. Each state is described by truth assignments to a set of propositions.

These propositions are as follows:

$S = \{s_C^1, \ldots, s_C^{|C|}\}$ - As before, a set of propositions, one for each agent indicating their support for C.

$V = \{v^0, \ldots, v^{|C|}\}$ - A set of propositions representing possible values of $q\#(h, C)$; these are $q\#(h, C) \in \{0, \ldots |C|\}$. The intended meaning is that if v^k then $q\#(h, C) = k$.

We now make some assumptions regarding the original knowledge of agent i, these are intuitive facts regarding $LFP(C)$.

As $q\#(h, C) \in \{0, \ldots |C|\}$ we assume that $q\#(h, C)$ may take exactly one value:

- $K_i(v^x \leftrightarrow \bigwedge_{y \neq x} \neg v^y)$

If $q\#(h, C) = |C|$ then agent h does not support (objects to) C:

- $K_i(v^{|C|} \leftrightarrow \neg s_C^h)$

Agent i knows the number of agents that support C.

To represent this we define a series of formulae representing an agent's knowledge of how many agents, including itself, support C. Let s_k be a formula defined for each k

in $\{0, \ldots, |C|\}$. Each formula is a disjunction of conjunctions of the form $s_C^{i_1} \wedge \cdots \wedge s_C^{i_k}$ and so each resembles the formula $s_{q\#(i,C)}$. Hence, the formula s_k holds in all possible worlds where at least k agents from C support C.

Worlds in which exactly k agents support C can be distinguished as those in which s_k holds but s_{k+1} does not.

Agent i knows that if agents h supports C then $q\#(h, C)$ is at most the number of other agents which also support C (from Definition 3.2). In possible worlds where agents h support C and exactly l other agents also support C then s_k holds where $k = l + 1$, it must be that $q\#(h, C) \leq l$, or equivalently $q\#(h, C) < k$. Therefore in these possible worlds it cannot be that v^y holds for $y \geq k$.

- $K_i(s_C^h \wedge (s_k \wedge \neg s_{k+1}) \leftrightarrow \bigwedge_{y \geq k} \neg v^y)$

Agents i know that if at the end of round m agent h has not announced s_C^h then $q\#(h, C)$ is strictly greater than the number of agents that had announced at the end of round $m - 1$.

- $\neg K_i^m(s_C^h) \wedge K_i^{m-1}(s_k \wedge \neg s_{k+1}) \leftrightarrow \bigwedge_{y \leq k} \neg v^y$

Agent i knows that if s_C^h is announced in round m then $q\#(h, C)$ is at most the number of agents that had announced at the end of round $m - 1$.

- $\neg K_i^{m-1}(s_C^h) \wedge K_i^m(s_C^h) \wedge K_i^{m-1}(s_k \wedge \neg s_{k+1}) \leftrightarrow \bigwedge_{y > k} \neg v^y$

To complete the proof, suppose that in rounds $m-2, m-1$ there are, respectively, a, b agents supporting C and that s_C^h is announced in round m.

At the end of round $m - 1$, $\neg K_i^{m-1}(s_C^h)$ hence i derives $\neg v^y$ for all $y \leq a$ thus all worlds where $q\#(h, C) \leq a$ are no longer considered possible.

At the end of round m, $K_i^m(s_C^h)$ hence i derives $\neg v^y$ for all $y > b$ thus all worlds where $q\#(h, C) > b$ are no longer considered possible.

Thus in all worlds considered possible by i the inequality $a < q\#(h, C) \leq b$ holds.

Since we assume that all original knowledge is common knowledge and all announcements are public it remains to observe that $a = |f_C^{m-2}(\emptyset)|$ and $b = |f_C^{m-1}(\emptyset)|$ and since s_C^h is announced in round m, $h \in f_C^m(\emptyset) \setminus f_C^{m-1}(\emptyset)$. □

In general then, agents may only know some inequality about the quorum values of other agents. However, special cases do exist where agents may acquire more precise information regarding the thresholds of other agents. Such revelations occur where agents make their announcement in the first round (in which case, clearly their threshold is zero) or where in the previous round only a single agent has announced s^i_C. Consider some round m where $|f^{m-2}_C(\emptyset)| - |f^{m-1}_C(\emptyset)| = 1$, in this case it will become common knowledge that $\forall i \in f^m_C(\emptyset) \setminus f^{m-1}_C(\emptyset), q\#(i, C) = |f^{m-1}_C(\emptyset)|$. The limiting case is found where in each round exactly one agent announces s^i_C in this event the exact value of $q\#(i, C)$ for each agent will become common knowledge.

3.8.2 Weak consensus

The $GFP(C)$ protocol requires only a single round and hence reveals less information regarding the threshold values of the agents. An agent announces o^i_C only if $q(i, C) > \frac{|C|-1}{|C|}$ therefore any agent making an announcement in the $GFP(C)$ protocol reveals that $q\#(i, C) > |C| - 1$. Similarly, any agent that has not announced o^i_C reveals that $0 \leq q\#(i, C) < |C|$.

3.9 Discussion and Related Work

Consensus games (CGs) are a novel approach to the problem of coalition formation in multi-agent systems. We have presented two solution concepts for CGs corresponding to the least and greatest fixed points of the function f_C. Our two solution concepts of strong and weak consensus represent different approaches to reaching consensus. Strong consensus requires that all agents support the formation of a coalition, whereas, weak consensus requires only that no agent objects to the formation of a coalition.

Since all strong consensus coalitions are weak consensus coalitions (but not vice versa) weak consensus coalitions may form where strong consensus coalitions may not. In both solution concepts objection, by even a single agent, is a barrier to coalition formation. For strong consensus it is necessary (but not sufficient) that at least one agent unconditionally supports the formation of a coalition; this requirement is relaxed in weak consensus. Where agent i unconditionally supports the formation of coalition C, $q(i, C) = 0$. Intuitively, we can say that agent i is 'particularly keen' that coalition C forms. Although weak consensus can be achieved in a broader range of situations this permissive quality may come at a cost. In particular, a weak consensus coalition may contain *no* agents that are particularity keen for that coalition to form.

Strong consensus resembles the threshold model of Granovetter [1978] since in that model at least one agent must 'initiate', by unconditionally supporting. Weak consensus resembles the threshold model presented in Chwe [1999]; since in that model an initiating agent is not necessary. CGs focus on the special case of consensus, an important aspect of decision making that has not been considered in previous work concerning threshold models. The problem of reaching consensus is particularly salient in a computation setting such as ours. Without a mechanism for consensus the coordination of distributed processes is all but impossible [Turek and Shasha, 1992].

CGs extend the threshold models of Granovetter [1978] and Chwe [1999] beyond binary choice decisions to the more general problem of coalition formation. In doing this we have developed notions of collective rationality and coalitional stability for threshold models. Such properties of threshold models have not been addressed in previous work.

The central proposition of CGs is that agents' choices are conditioned by the number of other agents also making some choice. This premise is found in the anonymous games of Daskalakis and Papadimitriou [2007]. In those games the individual utility of participation in some coalition is independent of the identities of the agents concerned; hence, factors including the size of the coalition become determinants of an agent's choice. In general, however, CGs are non-anonymous therefore, for example, an agent could object ($q(i, C) = 1$) to participation in any coalition in which some other, specific, agent participates.

The notion that an agent's behaviour may influence that of others is also found in imitation games [McLennan and Tourky, 2010]; in these, two player games, players assume the roles of leader and follower. Through the payoff structure the follower is motivated to act in consensus with the leader. Such a situation can be modelled as a CG where the leader has a threshold of zero, and the follower has a threshold of at most one half.

CGs have some similarities to Qualitative Coalitional Games (QCGs) [Wooldridge and Dunne, 2004]. It is therefore interesting to compare CGs and QCGs, especially with respect to the size of representation and the complexity of similar decision problems. A QCG Γ may be represented as an $(n+3)$ tuple $\Gamma = \langle \mathsf{G}, \mathsf{Ag}, \mathsf{G}_1 \ldots \mathsf{G}_n, \mathsf{V} \rangle$ where $\mathsf{G}_i \subseteq \mathsf{G}$ represents each agent's $\mathsf{i} \in \mathsf{Ag}$ set of goals and $\mathsf{V} : 2^{\mathsf{Ag}} \rightarrow 2^{2^{\mathsf{G}}}$ is the characteristic function of the game mapping each possible coalition of agents to the sets of goals that coalition can achieve. In QCGs:

- A set of goals $\mathsf{G}' \subseteq \mathsf{G}$ is *feasible* for a coalition $\mathsf{C} \subseteq \mathsf{Ag}$ if $\mathsf{G}' \in \mathsf{V}(\mathsf{C})$.
- A set of goals $\mathsf{G}' \subseteq \mathsf{G}$ *satisfies* an agent $\mathsf{i} \in \mathsf{C} \subseteq \mathsf{Ag}$ if $\mathsf{G}' \cap \mathsf{G}_i \neq \emptyset$.
- A coalition $\mathsf{C} \subseteq \mathsf{Ag}$ is *successful* if there exists some set of goals $\mathsf{G}' \subseteq \mathsf{G}$ such

that G' is feasible for C and G'satisfies at least all agents $\mathsf{i} \in \mathsf{C}$.

There is no immediate direct correspondence between goals in QCGs and quorum thresholds in CGs, apart from an intuition that the agent may have a lower quorum threshold for a coalition if it is more likely to achieve the agent's goals. However, the parallels between successful coalitions in QCGs and consensus coalitions in CGs are clearer: both capture the notion of it being rational for an agent to join a coalition. The worst case size of the game representation for QCGs is the characteristic function where each coalition can enforce any subset of goals. There are 2^n coalitions and 2^m subsets of goals, so the worst case size of V is $O(2^{n+m})$. For CGs, the representation is exponential in n. Complexity results for QCGs in [Wooldridge and Dunne, 2004] are given as a function of the size of representation, where the characteristic function is represented by a propositional formula Ψ (which as noted may be exponential in the number of agents and goals, but generally will be more concise than a naive representation of V). The successful coalition problem is NP in the size of the representation. It corresponds to our SCC problem which is linear in the number of agents (hence also in the size of representation).

The problem of verifying equilibria in CGs can be solved in time which is polynomial in the number of agents, whereas, the problem of verifying Nash equilibria in normal form game theoretic models requires time which is polynomial in the product of the number of players and of alternative strategies [Tadjouddine, 2007]. The complexity of computing the existence of Nash equilibria in normal form games is $PPAD$-complete [Chen and Deng, 2006]. $PPAD$ is a subclass of $TFNP$, the functional analogue of $NP \cap co\text{-}NP$ [Megiddo and Papadimitriou, 1991]. We show that the problem of computing existence of equilibria such as strong or weak consensus coalitions in CGs is in $NP(n)$. Our results suggest that the problem of computing the existence of equilibria in CGs is no more complex than it is in the typical game theoretic case.

Following the practices of game theoretic research [Harsanyi, 2004] we examined CGs under incomplete information; we defined and analysed protocols for both strong and weak consensus. Generally, the $LFP(C)$ protocol for strong consensus reveals more information than the $GFP(C)$ for weak consensus. This is not only true regarding the threshold values of the agents but is also true when considering the stability of the coalitions. Under strong consensus, it is possible for the agents to achieve common knowledge of the quality (number of rounds required for reaching consensus) of the coalition as they are running the $LFP(C)$ protocol. No such information is available in the case of the weak consensus protocol.

We have proven that both strong consensus and weak consensus can be established with a number of messages that is polynomial (linear) in the number of agents. That we find tractable results for reaching consensus under incomplete information is particularly promising; as Harsanyi [2004] observes, in many real-world situations participants often lack full information. Our results suggest that practical systems using our protocols are achievable.

Chapter 4

Consensus Action Games

4.1 Consensus Action Games

Consensus action games extend consensus games by the inclusion of a representation of the joint action space. Our motivation for this extension follows from the generally accepted view that agents are capable of performing actions. One common model of possible outcomes for joint activity is to define a joint action space as the Cartesian product of individual action sets (of each agent). This is reminiscent of the strategy space found in the normal form games of non-cooperative game theoretic models. Although such an approach is appropriate for describing abstract strategy spaces it presents certain difficulties when considering concrete action. Notably, for situated agents in domains where their actions require external resources which are in some way bounded. Bounded resource may constrain the ways in which agents act together. For example, prior to use, a resource may require that an exclusive lock is obtained. Whilst it may be that any agent could perform an action requiring the resource it is necessary that the agents' actions are coordinated. The agents must act in consensus, such that no two agents simultaneously attempt to utilise the resource. Such constraints are not easily captured by models where the joint action space is unrestricted.

Instead, to provide for the modelling of such real-world situations we will introduce the set Ac representing the space of all possible individual actions and the structure J which holds an explicit representation of only those joint actions permitted according to the domain of interest. Formally:

Definition 4.1. *A consensus action game is a tuple* $\Gamma = \langle A, Ac, J, q \rangle$ *where:*

A *is a finite set of agents,* $\{1, \ldots, n\}, n \geq 2$.

Ac is a finite, non empty set of possible actions $\{1, \ldots, m\}$.

J is a set of joint actions; each joint action is a set of pairs (i, a), where $i \in A$ and $a \in Ac$, specifying the action performed by each agent participating in the joint action. Each participating agent performs exactly one action. The set $J_i = \{j \in J \mid (i, a) \in j\}$ indicates the set of joint actions in which agent i may participate; $J_C = \{j \in J \mid \{i \mid (i, a) \in j\} = C\}$ indicates the set of all joint actions that can be performed by the set of agents $C \subseteq A$.

q is a quorum function. The partial function $q : \{(i, j) \mid i \in A, j \in J_i\} \rightarrow [0, 1]$ takes an agent $i \in A$ and a joint action j in J_i and returns a number in the interval [0,1].

In a similar manner to CGs, the value of the quorum function in CAGs indicates the agent's degree of support, however, in CAGs agents do not explicitly express their support for the formation of coalitions. Instead, agents' support is for a joint action performed by a coalition. Thus: for an agent $i \in C \subseteq A$ and joint action $j \in J_C$, the quorum function $q(i, j)$ gives the minimum proportion of agents in C which must support j in order that i will support j.

Individually rational behaviour in CAGs is exhibited as follows: where $q(i, j) = 0$ agent i shows unconditional support for j, where $0 < q(i, j) \leq \frac{|C|-1}{|C|}$ the agent shows conditional support for j; where $\frac{|C|-1}{|C|} < q(i, j) \leq 1$ the agent objects to j.

4.2 Consensus in CAGs

In this section we extend the notions of strong and weak consensus developed for CGs to accommodate the inclusion of joint actions introduced by CAGs. Following the notational conventions introduced in Chapter 3: for some agent $i \in C \subseteq A$ and some joint action $j \in J_C$ let $q\#(i, j)$ denote the number of other agents from C required to support j in order that i supports j; and let $n_k(j)$ denote the number of agents $i \in C$ with $q\#(i, j) = k$.

In CGs agents form coalitions by consensus. The analogue of these consensus coalitions in CAGs are *consensus actions*. Similarly, notions of strong and weak consensus coalitions found in CGs are represented by *strong consensus actions* and *weak consensus actions*, respectively:

Definition 4.2. *A joint action $j \in J$ is a strong consensus action for agents $C \subseteq A$ if:*

- $j \in J_C$
- $n_0(j) \neq 0$
- *if* $n_k(j) \neq 0$, *then* $\sum_{p<k} n_p(j) \geq k$

Definition 4.3. *A joint action* $j \in J$ *is a weak consensus action for agents* $C \subseteq A$ *if:*

- $j \in J_C$
- $\neg\exists i \in C$ *such that* $q(i, j) > \frac{|C|-1}{|C|}$

As for CGs, in CAGs these notions of consensus may also be formalised as greatest/least fixed point constructions. Consider the function $f_j : 2^A \to 2^A$ defined relative to $j \in J$, for some joint action $j \in J_C$ performed by agents $C \subseteq A$:

$$i \in f_j(Q) \text{ iff } i \in C \text{ and } |Q \cap C \setminus \{i\}| \geq q(i, j) \times |C|$$

A joint action $j \in J_C$ is a strong consensus action if C is the least fixed point of f_j. A joint action $j \in J_C$ is a weak consensus action if C is the greatest fixed point of f_j.

Consider the following example in which we will denote that agent $i \in A$ performs action $a \in A$ as i^a.

Example 4.1. *Newly engaged, Alice and Bob decide to cohabitate. Bob invites Alice and their new house-mates, Charles and Daphne, out for dinner. Charles is keen that they all go out to dine* (d)*. Alice is willing to join Bob and Charles, but only if Daphne also accompanies them. It is possible that all four will go out to dinner. However, if Daphne stays at home* (s) *Alice will stay with her. Still, Bob may dine together with Charles if Charles remains interested; Charles feels likewise. Happily, Daphne is delighted to be invited. We can formalise this example as the consensus action game* $\Gamma_3 = \langle A, Ac, J, q \rangle$ *where:*

$A = \{Al, Bo, Ch, Da\}$
$Ac = \{d, s\}$
$J = \{j_1, j_2, j_3\}$ *where:*
$j_1 = \{Al^d, Bo^d, Ch^d, Da^d\}$,
$j_2 = \{Bo^d, Ch^d\}$,
$j_3 = \{Al^s, Da^s\}$,

$$q(i,j) = \begin{cases} 0 & \textit{if } i \in \{Bo, Ch, Da\} \textit{ and } j = j_1 \\ 0.75 & \textit{if } i = Al \textit{ and } j = j_1 \\ 0.5 & \textit{if } i \in \{Bo, Ch\} \textit{ and } j = j_2 \\ 0.5 & \textit{if } i = Al \textit{ and } j = j_3 \\ 1 & \textit{if } i = Da \textit{ and } j = j_3 \end{cases}$$

In Example 4.1 j_1 is unconditionally supported by Bob, Charles and Daphne whilst Alice conditionally supports j_1 requiring at least three quarters of those involved to support before she will also support. The joint action j_1 in which all four go out to dine is a strong (and hence also weak) consensus action. Bob and Charles' position regarding dining without Alice and Daphne resembles that found in Example 3.2, thus j_2 is a weak consensus action. Finally, j_3 is not a consensus action as Daphne objects to staying at home when she could go out to dinner with the others.

4.3 Collective Rationality in CAGs

In Chapter 3 notions of collective rationality for CGs were introduced. These are different for strong and weak consensus coalitions. We will now revisit these definitions and consider their appropriateness for the CAG model.

For strong consensus the number of rounds required to establish consensus is associated with the effort required by the agents to reach consensus. This natural measure of the stability of a coalition is captured in the notion of the q-minimal core of a CG. The strong consensus coalition $C \subseteq A$ is in the q-minimal core of a CG only if there does not exist some other strong consensus coalition $C' \subset C$ that can reach consensus in strictly fewer rounds than required for C.

In CAGs each coalition may have multiple strong consensus actions, a situation which does not arise in CGs as, for CGs, each coalition may have at most one consensual outcome: the coalition is a strong consensus coalition. The q-minimal core as defined for CGs, where only proper subsets of a coalition may undermine a coalition's stability is therefore inappropriate for the CAG model as, although some $j \in J_C$ may be a strong consensus action there may also exist some other joint action $j' \in J_C$ where $rounds(j') < rounds(j)$. The existence of j' undermines the stability of j. A definition of collective rationality for strong consensus actions in CAGs must therefore respect this distinction.

In CAGs the analogue for the q-minimal core in CGs is the q_C*-minimal core* defined as containing only those joint actions for which the number of agents and the number of rounds required for strong consensus are minimal.

Definition 4.4. *For agents in $C \subseteq A$ the joint action $j \in J_C$ is a q_C-minimal consensus action only if:*

- *j is a strong consensus action,*
- *there exists no other strong consensus action $j' \in J_{C'}, j' \neq j$ for $C' \subseteq C$ where the number of rounds required to reach consensus for j' is strictly less than the number of rounds required to reach consensus regarding j.*

Formulation of the stable set for weak consensus in CAGs is more straightforward. Recall that in CGs a coalition is in the C-minimal core only if it is a weak consensus coalition and there does not exist some proper subset of that coalition which is also a weak consensus coalition. We formulate the notion of collective rationality for weak consensus in CAGs in a corresponding manner.

Definition 4.5. *For agents in $C \subseteq A$ the joint action $j \in J_C$ is a C-minimal consensus action only if:*

- *j is a weak consensus action,*
- *there exists no other weak consensus action $j' \in J_{C'}$ such that $C' \subset C$.*

The C-minimal core of a CAG aggregates weak consensus actions. Note, that, as in CGs we require that C' is a proper subset of C ; this implies that for a given coalition there may be several joint actions in the C-minimal core of a CAG.

4.4 Computational Complexity of CAGs

The characterisation of the computational complexity of consensus action games extends the characterisation for consensus games introduced in section 3.5. For purposes of comparison, decision problems regarding verification, existence and non-existence of consensus actions are again considered. In CGs each coalition has exactly one choice; in CAGs the coalition $C \subseteq A$ may have at most $m^{|C|}$ choices. It is interesting therefore to enquire regarding possible outcomes for some subset of the agents. A further type of decision problem, denoted GA, is introduced for this purpose. Below, decision problems of the following types are considered:

Consensus Action (CA): is a joint action a consensus action?

Consensus Action Exists (AE): does some group of agents have a consensus action?

No Consensus Action (NA): is it the case that no group of agents has a consensus action?

Group Consensus Action (GA): does a particular group of agents have a consensus action?

4.4.1 Representation

Here we consider the size and representational scheme for CAGs to be used as the input to these decision problems. Given a set of agents of size n and a set of actions of size m, in the worst case (when every set of agents can jointly execute any possible combination of actions) the set of joint actions J has cardinality $O(m^n)$, i.e., exponential in the number of agents.

The input to the decision problems, R, is an encoding of the quorum function q which must describe a threshold value for each agent of each joint action $j \in J$. This can be represented as a set of triples. Let $Ag : J \to \mathcal{P}(A)$ be a function that for a given joint action $j \in J_C$ returns $C \subseteq A$, the set of agents that may participate in joint action j. Then, for an agent $i \in C \subseteq A$ and a joint action $j \in J_C$ the triple $(i, j, q(i, j))$ gives the threshold of agent i for joint action j. The joint action j can be fully described by the set of triples $rep(j) = \{(i, j, q(i, j)) \mid i \in Ag(j), q(i, j) \in [0, 1]\}$; the overall representation of the game is then the set $R = \{rep(j) \mid j \in J\}$.

In the following analyses it is assumed that the function Ag runs in at most $O(n)$. In addition we assume that R is implemented as a random access data structure and that the size (number of elements) in R can be determined in $O(log|J|)$. Hence, the following results are given for the non-deterministic random access machine (NRAM) model of computation.

4.4.2 Complexity of strong consensus decision problems

STRONG CONSENSUS ACTION(SCA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$ and a joint action $j \in J$, is j a strong consensus action?

A deterministic algorithm must verify that $Ag(j)$ is the least fixed point of f_j. Algorithm 4.1 runs in time linear in n; the problem therefore has time complexity $O(n)$ and therefore lies within $P(n)$. The algorithm is similar to Algorithm 3.1 with only slight modification to accommodate the representation of joint actions in CAGs as opposed to coalitions in CGs.

Algorithm 4.1 Is j a strong consensus action.

```
function SCC(R, j)
    C ← Ag(j)
    array support[|C| + 1] ← {0, ..., 0}
    for all (i, j, q) ∈ j do
        k ← ⌈q × |C|⌉
        support[k] ← support[k] + 1
    end for
    s ← support[0]
    for k from 1 to |C| do
        if k ≤ s then
            s ← s + support[k]
        else
            return false
        end if
    end for
    return true
end function
```

STRONG CONSENSUS ACTION EXISTS (SAE)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$ is there some $j \in J$ that is a strong consensus action?

A deterministic algorithm would iterate through each $j \in J$ checking if j is a strong consensus action, returning yes if a strong consensus action is found and no, otherwise. The problem therefore has time complexity of $O(n \times |J|)$.

A non-deterministic algorithm first guesses an index of an action j in J (this can be done in $O(log(|J|)) \leq O(n)$ by the assumption that the size of J is obtainable in $O(log(|J|))$, and then checks j is a strong consensus action. This verification can be done in time linear in n using Algorithm 4.1. This gives a non-deterministic linear time algorithm for a random access machine. Hence, the problem is in $NP(n)$ for NRAM.

NO STRONG CONSENSUS ACTION (SNA)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$, is it the case that no $j \in J$ is a strong consensus action?

A deterministic algorithm would iterate through each $j \in J$ checking if j is a strong consensus action, returning no if a strong consensus action is found and yes, otherwise. The problem therefore also has time complexity of $O(n \times |J|)$.

Since the problem of the existence of a quorum consensus action is in $NP(n)$ (from SAE, above) its compliment, SNA, is in *co*-$NP(n)$ for NRAM.

STRONG CONSENSUS GROUP ACTION (SGA)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$ and a subset of agents $C \subseteq A$, is there a quorum consensus action for C?

A deterministic algorithm must verify that $\exists j \in J_C$ such that C is the least fixed point of f_j. The problem therefore has time complexity of $O(n \times |J|)$

A non-deterministic algorithm first guesses an index of an action $j \in J$ and then checks that $Ag(j) = C$ and that j is a strong consensus action. This gives a non-deterministic linear time algorithm, hence the problem is in $NP(n)$ for NRAM.

q_C-MINIMAL STRONG CONSENSUS ACTION (QM-SCA)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$, and a joint action $j \in J$, is j a q_C-minimal consensus action?

Let $J_{C' \subseteq C} = \{j \in J \mid \{i \mid (i, a) \in j\} \subseteq C\}$ denote the set of all joint actions for all groups of agents $C' \subseteq C$. A deterministic algorithm must verify that j is a strong consensus action and that $\neg\exists j' \in J_{C' \subseteq C}, j \neq j'$ such that j' is also a strong consensus action and $rounds(j') < rounds(j)$. For a strong consensus action $j \in J$ the function $rounds : J \to \mathbb{N}$ gives the number of rounds necessary for there to be strong consensus about j; this function can be implemented in a manner similar to algorithm 3.2 and therefore has time complexity $O(n)$. A deterministic algorithm for this problem therefore has time complexity $O(n \times |J|)$.

A non-deterministic algorithm for deciding that j is *not* a q_C-minimal consensus action first checks whether it is a strong consensus action (and returns yes if it is not) and if it is, compute $rounds(j)$ and guess a joint action $j' \in J$ and verify that $j' \neq j$, $Ag(j') \subseteq Ag(j)$ and $rounds(j') < rounds(j)$. The problem of deciding that j is *not* a q_C-minimal consensus action is therefore in $NP(n)$ on NRAM. Hence deciding whether j is a q_C-minimal consensus action is in *co*-$NP(n)$ on NRAM.

q_C-MINIMAL STRONG CONSENSUS ACTION EXISTS (QM-SAE)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$ is there some $j \in J$ that is a q_C-minimal consensus action?

If there exists some joint action $j \in J_C$ which is a strong consensus action then either that action, or an alternative joint action $j' \in J_{C' \subseteq C}, j' \neq j$ will be a q_C-minimal consensus action. Therefore this decision problem can be solved in a manner similar to SAE. Hence, the problem is in $NP(n)$ for NRAM.

NO q_C-MINIMAL STRONG CONSENSUS ACTION (QM-SNA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$ is there no $j \in J$ that is a q-minimal consensus action?

If there exists some joint action $j \in J_C$ which is a strong consensus action then either that action, or an alternative joint action $j' \in J_{C' \subseteq C}, j' \neq j$ will be a q_C-minimal consensus action. Therefore this decision problem can be solved in a manner similar to SNA. Hence, the problem is is in *co*-$NP(n)$ for NRAM.

q_C-MINIMAL STRONG CONSENSUS GROUP ACTION (QM-SGA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, and a subset of agents $C \subseteq A$ is there a q_C-minimal consensus action for C?

A deterministic algorithm must verify that $\exists j \in J_C$ such that j is a strong consensus action and that $\neg \exists j' \in J_{C' \subseteq C}, j \neq j'$ such that j' is also a strong consensus action where consensus is reached in strictly fewer rounds for j' than for j. Therefore the problem has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first calls an $NP(n)$ oracle to check that C has a strong consensus action; if there is a strong consensus action for C it then calls an $NP(n)$ oracle to check whether any $C' \subseteq C$ has a strong consensus action converging in strictly fewer rounds. Hence the problem is in $D^p(n)$ for NRAM. The Difference class D^p is the class of problems which are in the difference of two NP classes of problems [Papadimitriou, 1994].

4.4.3 Complexity of weak consensus decision problems

In this section decision problems for weak consensus actions and C-minimal consensus actions are considered.

WEAK CONSENSUS ACTION (WCA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, and a joint action $j \in J$ is j a weak consensus action?

A deterministic algorithm must verify that $\neg \exists i \in Ag(j)$ such that $q(i, j) > \frac{|Ag(j)|-1}{|Ag(j)|}$. This can be achieved by iteration over $rep(j)$ checking that $q(i, j) \leq \frac{|Ag(j)|-1}{|Ag(j)|}$. Hence, the problem can be decided in time that is linear in the number of agents and is in $O(n)$.

WEAK CONSENSUS ACTION EXISTS (WAE)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, is there some $j \in J$ that is a weak consensus action?

A deterministic algorithm must verify that $\exists j \in J$ such that $\neg \exists i \in Ag(j)$ such that $q(i, j) > \frac{|Ag(j)|-1}{|Ag(j)|}$. Hence the problem has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first guesses an index of a joint action $j \in J$ and then checks that j is a weak consensus action. $O(log(|J|)) \leq O(n)$ operations are required to guess an index, whilst verification that j is a weak consensus action can be performed in $O(n)$ (from WCA, above). This gives a non-deterministic random access algorithm, hence the problem is in $NP(n)$ for NRAM.

NO WEAK CONSENSUS ACTION (WNA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, is it the case that no $j \in J$ is a weak consensus action?

A deterministic algorithm must verify that $\neg \exists j \in J$ such that j is a weak consensus action. Hence the problem has time complexity $O(n \times |J|)$

The problem of determining if there exists some weak consensus action, WAE is in $NP(n)$; the complement of that problem is WNA. Hence, this problem is in *co*-NP for NRAM.

WEAK CONSENSUS GROUP ACTION (WGA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, and a subset of agents $C \subset A$ is there a weak consensus action for C?

A deterministic algorithm must verify that $\exists j \in J$ such that $Ag(j) = C$ and j is a weak consensus action. Hence the problem has time complexity $O(n \times |J|)$.

A non deterministic algorithm first guesses an index of a joint action $j \in J$ and then checks that $Ag(j) = C$ and that j is a weak consensus action. This gives a non-deterministic linear time algorithm for a random access machine. Hence the problem is in $NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS ACTION (CM-WCA)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$, and a joint action $j \in J$ is j a C-minimal consensus action?

Let $J_{C' \subset C} = \{j \in J \mid \{i \mid (i, a) \in j\} \subset C\}$ denote the set of all joint actions for all groups of agents $C' \subset C$. A deterministic algorithm must verify that j is a weak consensus action and that $\neg\exists j' \in J_{C' \subset C}$ such that j' is also a weak consensus action. This gives a deterministic algorithm with time complexity $O(n \times |J|)$.

A non-deterministic algorithm to solve the complement of this problem (decide whether a joint action is *not* a C-minimal consensus action) first checks whether j is a weak consensus action (and returns true if it is not); if j is a weak consensus action, it will guess an index of an action $j' \in J$ and check that $Ag(j') \subseteq A$ and that j' is a weak consensus action. So the problem of deciding whether an action is *not* a C-minimal weak consensus action is in $NP(n)$ on NRAM. Hence deciding whether an action is a C-minimal consensus action is in *co*-$NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS ACTION EXISTS (CM-WAE)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$, is there some $j \in J$ that is a C-minimal consensus action?

If there exists some joint action j which is a weak consensus action then either j itself or some other joint action for a proper subset of $Ag(j)$ will be a C-minimal weak consensus action. Therefore this problem can be solved in a manner similar to WAE; the problem therefore has time complexity $O(n \times |J|)$ and is in $NP(n)$ for NRAM.

NO C-MINIMAL WEAK CONSENSUS ACTION (CM-WNA)

Given a CAG $\Gamma = \langle A, Ac, J, q\rangle$, is it the case that no $j \in J$ is a C-minimal consensus action?

If $j \in J_C$ is a weak consensus action then either j or some $j' \in J_{C' \subset C}$ is a C-minimal consensus action. Therefore this problem can be solved in a similar manner to WNA and has time complexity $O(n \times |J|)$.

From CM-WAE (above) the solution to the complement of this problem is in $NP(n)$; hence, this problem is in $co\text{-}NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS GROUP ACTION (CM-WGA)

Given a CAG $\Gamma = \langle A, Ac, J, q \rangle$, and a subset of agents $C \subseteq A$ is there a C-minimal consensus action for C?

A deterministic algorithm must verify that $\exists j \in J_C$ such that j is a weak consensus action and that $\neg \exists j' \in J_{C' \subset C}$ such that j' is also a weak consensus action. Therefore the problem has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first calls an $NP(n)$ oracle to check that C has a weak consensus action; if there is a weak consensus action for C it then calls an $NP(n)$ oracle to check whether any $C' \subset C$ has a weak consensus action. Hence the problem is in $D^p(n)$ for NRAM.

4.5 Comparison of CGs and CAGs

The model for CGs contains only two elements, a finite set of agents, A, and the quorum function q. CAGs extend CGs by introducing a further two elements, a set of actions, Ac and a set of joint actions, J. Each joint action $j \in J$ for the subset of agents $C \subseteq A, C = Ag(j)$ ascribes a single action $a \in Ac$ to each agent $i \in C$.

The solution concepts of strong and weak consensus coalitions in CGs can be translated into CAGs as strong and weak consensus actions; these are joint actions in which it is individually rational (under the prevailing solution concept) for agents to participate. The solution concept for collective rationality of weak consensus coalitions, the C-minimal core was also unaffected by the introduction of joint actions to the model. Interestingly, however, the solution concept of q-minimal consensus in CGs has been found to be inappropriate for CAGs.

The q-minimal core of a CG aggregates strong consensus coalitions where no subset of agents in a coalition can reach consensus in fewer rounds by defection to a smaller coalition; hence, such coalitions are collectively rational. However, the addition of joint actions in CAGs significantly alters the structure of outcomes in the games, from a single consensus outcome per coalition in CGs to potentially multiple consensus outcomes in CAGs. Consequently the question of whether or not a joint action (rather than a coalition) is collectively rational for its associated coalition of agents is important for CAGs.

It does not appear collectively rational for some subset of agents $C \subseteq A$ to perform the strong consensus action $j \in J_C$ if there exists some alternative strong consensus action $j' \in J_C, j \neq j'$ such that j' requires fewer consensus rounds than j. Although the formation of the coalition C is a stable outcome, the joint action j is not. The q_C-minimal core for CAGs thus extends the q-minimal core for CGs. The q-minimal core considers only proper subsets of coalitions, whereas the q_C-minimal core also considers the stability of joint action within the coalition itself.

The impact of the introduction of joint actions to the model is evident in the complexity results for decision problems. Table 4.1 summarises the complexity results for CGs and CAGs. Encouragingly, computing whether a joint action is a consensus action (strong or weak) can be accomplished in time that is linear in n, the number of agents. Hence, both SCA and WCA are in $P(n)$, as are their counterparts SCC and WCC for CGs. This suggests that knowledge based protocols for strong and weak consensus in CAGs should be no more complex than those given for CGs in section 3.7.

Table 4.1: Comparison of complexity results for CGs and CAGs.

Consensus Games			**Consensus Action Games**		
Decision Problem	Time Complexity	Complexity Class	Decision Problem	Time Complexity	Complexity Class
SCC	$O(n)$	$P(n)$	SCA	$O(n)$	$P(n)$
SCE	$O(n2^n)$	$NP(n)$	SAE	$O(n \times \lvert J \rvert)$	$NP(n)$
SNC	$O(n2^n)$	*co*-$NP(n)$	SNA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
–			SGA	$O(n \times \lvert J \rvert)$	$NP(n)$
QM-SCC	$O(n2^n)$	*co*-$NP(n)$	QM-SCA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
QM-SCE	$O(n2^n)$	$NP(n)$	QM-SAE	$O(n \times \lvert J \rvert)$	$NP(n)$
QM-SNC	$O(n2^n)$	*co*-$NP(n)$	QM-SNA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
–			QM-SGA	$O(n \times \lvert J \rvert)$	$D^p(n)$
WCC	$O(n)$	$P(n)$	WCA	$O(n)$	$P(n)$
WCE	$O(n2^n)$	$NP(n)$	WAE	$O(n \times \lvert J \rvert)$	$NP(n)$
WNC	$O(n2^n)$	*co*-$NP(n)$	WNA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
–			WGA	$O(n \times \lvert J \rvert)$	$NP(n)$
CM-WCC	$O(n2^n)$	*co*-$NP(n)$	CM-WCA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
CM-WCE	$O(n2^n)$	$NP(n)$	CM-WAE	$O(n \times \lvert J \rvert)$	$NP(n)$
CM-WNC	$O(n2^n)$	*co*-$NP(n)$	CM-WNA	$O(n \times \lvert J \rvert)$	*co*-$NP(n)$
–			CM-WGA	$O(n \times \lvert J \rvert)$	$D^p(n)$

For problems concerned with existence, verification or collective rationality it is, in the worst case, necessary that a deterministic algorithm process the entirety of the input representation, R. Hence, many decision problems for CGs have deterministic solutions with time complexity that is exponential in the number of agents: $O(n2^n)$. The time complexity of comparable problems for CAGs has been found to be $O(n \times |J|)$. This change is merely a consequence of the structural difference between CGs and CAGs in the way that the outcome space is represented. In CGs, where agents form coalitions, the number of possible coalitions, and hence possible outcomes is exponential in the number of agents; there must be precisely $2^n - 1$ (the empty coalition is disregarded) outcomes. For CAGs the size of the outcome space is not a function of n, it is a function of the cardinality of the joint action set, J. This is quite arbitrary; therefore, whilst the worst case upper bound on $|J|$ is $O(m^n)$ (hence, still exponential in the number of agents), for any particular CAG $|J|$ may be significantly smaller than this (as is the case for Example 4.1). Therefore decision problems for CAGs are no more or less tractable than they are for CGs; indeed those which are in (*co*-)NP for CGs have remained in (*co*-)NP, respectively.

In addition to considering decision problems previously identified for CGs a new type of decision problem (GA) has been introduced. Problems of this type are concerned with the existence of stable outcomes for a given subset of agents. The problems SGA and WGA were found to have similar complexity to their more general counterparts SAE and WAE. In CAGs, for these problems at least, problems concerning a specific coalition are no easier than those considering any coalition. Note that this is not the case for the CGs model; the problems SCC and WCC, is $C \subseteq A$ a strong/weak (resp.) consensus coalition, were found to be in $P(n)$. Whereas the corresponding problems for any coalition, SCE and WCE, are in $NP(n)$.

The problems QM-SGA and CM-WGA enquire regarding the existence of collectively rational outcomes for a known subgroup of agents. These problems were found to be in D^p. The difference class D^p contains problems which are solvable using two NP oracles where the first oracle answers "yes" and the second, "no". For example the problem SAT-UNSAT (given two formulae is it the case that the first is satisfiable and the second is not) is D^p-complete [Papadimitriou and Wolfe, 1985]. Wooldridge and Dunne [2004] show that for QCG's decision problems similar to QM-SGA and CM-WGA (such as minimal successful coalition) are also D^p-$complete$.

4.6 Discussion and Related Work

In this chapter we have described consensus action games (CAGs); these extended consensus games by the inclusion of joint actions. Given the coalition of agents, $C \subseteq A$, and set of individual actions, Ac, the most straightforward representation of the joint action space is the Cartesian product of these sets. Such a joint action space would contain all possible combinations of agents $i \in C$ and actions $a \in Ac$. We chose to eschew this naive, implicit representation of the joint action space. Rather, in the CAG model we have an *explicit* set of joint actions, J.

The coalition of agents, $C \subseteq A$ may only perform those joint actions found in the set $J_C \subseteq J$. Each joint action, $j \in J_C$ ascribes exactly one individual action $a \in Ac$ to each participating agent $i \in C$. This approach to modelling joint actions does not preclude the straightforward case but also permits CAGs to capture subtle facets of joint action in the real-world.

In particular, it is not necessarily the case that all agents have the capability to perform the exact same range of individual actions. Certain agents may possess abilities which are specific to only themselves or a subset of agents. For example not all agents may posses the facility to transport a large or heavy item. Our representation of the joint action space is well suited to modelling situations where agents are heterogeneous with respect to their individual actions. Such distinction is not possible under the naive representation.

Even where agents' abilities are homogeneous it is not necessarily the case that all agents could perform an identical action simultaneously. The space of joint action is frequently constrained by the physical laws governing action in the real world. Consider, for example, a domain in which chairs are suitable only for one occupant and there exists but a single chair. Even though each agent may be capable of sitting in the chair it is not possible that they all do so concurrently; this cannot be captured by the naive representation. Our explicit model of joint action in CAGs accommodates situations where agents' actions utilise resources and respects that, in reality, resources are bounded.

The majority of prior work concerning joint action in multi-agent systems has tended to cast the problem of collaborative joint action as a team planning problem. The SharedPlans of Grosz and Kraus [1996, 1999]; Grosz and Sidner [1990], joint intentions of Cohen and Levesque [1991]; Levesque et al. [1990] and STEAM model of teamwork [Tambe, 1997] all view the joint action problem in this way. In these approaches the focus has been on the specification of a plan of joint actions in pursuit of a shared goal. Typically the agents are assumed to be benevolent, therefore, the matter of whether or not the agents will agree to act jointly has not been of great importance.

In contrast to those team-centric views Shehory and Kraus [1998] describe a method for task allocation in multi-agent systems in which agreement is central. In their model agents may gain utility by the completion of tasks; to be completed successfully each task requires a set of capabilities. Typically each agent is endowed with only certain capabilities, hence agents are motivated to form coalitions in order to achieve tasks.

The notion of agents' capabilities found in Shehory and Kraus [1998] can be associated with our notion of individual actions in CAGs. The set of capabilities required to complete a task resembles our notion of joint action in that both convey the idea of a set of agents performing a collective act. However, although Shehory and Kraus specify the actions that should be performed (such that a task is completed) their model does not capture which agent will perform which action. Moreover, Shehory and Kraus consider coalition formation and joint action from a purely economic (game theoretic) perspective. Ours is the first work to consider consensual joint action in multi-agent systems where coalitions of agents are formed and joint actions are selected according to endogenous thresholds.

Chapter 5

Consensus Action Games with Goals

5.1 Consensus Action Games with Goals

Typically agents are characterised as performing actions in pursuit of goals. Consensus action games with goals (CAG-Gs) extend consensus action games by including representations of the agents' motivational state. The motivations of an agent in CAG-Gs are represented by the goals of the agent and an ordering over these goals. Each agent would like to achieve one or more of their goals, these can be achieved only through participation in consensus joint actions. The inclusion of information concerning the motivations of the agents permits us to investigate rational specification of the quorum function for CAG-Gs.

CAG-Gs model a very general notion of agents' goals; we make no assumptions regarding the nature of the goals that agents may have. We do not, for example, assume that agent's goals are necessarily compatible with one another. Therefore CAG-Gs are suitable for modelling circumstances in which the agents have heterogeneous and possibly conflicting goals. CAG-Gs also model a very general notion of ordering over goals. In CAG-Gs agents have orderings over subsets of their goals. This allows for agents to achieve multiple goals where some combinations of goals are more (or less) desirable than others. For example given the goals 'to marry Mary' and 'to marry Sue' it is better to marry Mary or Sue than it is to marry Mary *and* Sue. Therefore we do not assume that monotonically increasing (by membership) sets of goals are increasingly desirable.

In CAG-Gs each agent has a preorder (transitive and reflexive) over subsets of their goals. This contrasts with the approach of social choice theory where, typically, agents' orderings are partial (transitive, reflexive and asymmetric). We do not require asymmetry as, for two sets of goals, it may be that these are ranked equally in the ordering; however this does not imply that these are the same set of goals. The orderings over the agents' goals are not required to be total (defined for every possible pair). Thus in CAG-Gs, agents may possess goals which are incomparable. This differentiates CAG-Gs from other common models including preference orderings. We see this difference as beneficial for many real-world settings where goals may not necessarily be comparable and not all equally preferred things are the same.

Whilst the introduction of goals and orderings over subsets of goals has implications for how quorum values may be specified, the mechanisms by which agents may reach consensus are not affected by the inclusion of these elements in the model. Consequently, definitions for strong and weak consensus, together with their associated core concepts for CAG-Gs are the same as for CAGs. The definitions given in Chapter 4 are not repeated here.

Definition 5.1. *A consensus action game with goals (CAG-G) is a tuple* $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ *where:*

A *is a finite set of agents,* $\{1, \ldots, n\}, n \geq 2$.

Ac *is a finite, non empty set of possible actions* $\{1, \ldots, m\}$.

J *is a set of joint actions; each joint action is a set of pairs* (i, a)*, where* $i \in A$ *and* $a \in Ac$*, specifying the action performed by each agent participating in the joint action. Each participating agent performs exactly one action. The set* $J_i = \{j \in J \mid (i, a) \in j\}$ *indicates the set of joint actions in which agent* i *may participate;* $J_C = \{j \in J \mid \{i \mid (i, a) \in j\} = C\}$ *indicates the set of all joint actions that can be performed by the set of agents* $C \subseteq A$.

G *is a finite, non empty set of possible goals.*

$goals$ *is a function* $goals : A \longrightarrow 2^G$*. For each agent* $i \in A$*,* $goals(i) \subseteq G$ *gives the set of goals agent* i *would like to achieve.*

$achieve$ *is a function* $achieve : J \longrightarrow 2^G$*. For each joint action* $j \in J$*,* $achieve(j) \subseteq G$ *gives the set of goals achieved through the performance of* j*. We write:*

$achieve_i(j) = achieve(j) \cap goals(i)$ *to denote the set of goals achieved by agent* $i \in A$ *through participation in joint action* $j \in J_i$.

$\succeq_i$ *is a preorder (transitive and reflexive)* $\succeq_i \subseteq 2^{goals(i)} \times 2^{goals(i)}$. *For each agent* $i \in A$, $\succeq_i$, *is an ordering over subsets of goals. For the sets of goals* $g, g' \in \mathcal{P}(goals(i))$ *we write:*

$g \succeq_i g'$ *to denote that for agent* i *the set of goals* g *is ranked at least as high as the set of goals* g',

$g \cong_i g'$ *for* $g \succeq_i g'$ *and* $g' \succeq_i g$ *to denote that agent* i *is indifferent between the sets of goals* g, g',

$g \succ_i g'$ *for* $g \succeq_i g'$ *and not* $g' \cong_i g$ *to denote that for agent* i *the set of goals* g *is ranked strictly higher than the set of goals* g'

We require that for $g, g' \in \mathcal{P}(goals(i)), g \neq g'$ *where* $g' = \emptyset$, $g \succ_i g'$. *Thus, achieving something is strictly better than achieving nothing.*

q *is a quorum function. The partial function* $q : \{(i,j) \mid i \in A, j \in J_i\} \rightarrow [0,1]$ *takes an agent* $i \in A$ *and a joint action* j *in* J_i *and returns a number in the interval [0,1].*

5.2 Rational Specification of q for CAG-Gs

In previous chapters we have relied upon an intuitive notion that quorum values represent the degree of support of an agent to participate in a coalition or joint action. The basis of this intuition is that: for a given agent $i \in A$ the fewer the number of other agents required to support a collective act before i also supports the more keen i is to participate in that act. Thus, in general, it is the case that agents are more keen to participate in collective acts where their threshold is lower.

In this section we are going to formalise our intuitive understanding of quorum values into a series of rationality postulates. These postulates, to which all CAG-Gs must adhere, constrain the specification of the quorum function in consensus action games with goals to only those functions which may be considered rational given the motivational state of the agents.

We begin by considering those joint actions the agents are most motivated to participate in. The motivation of agent $i \in A$ for participation in joint action $j \in J_i$ is that, through participation in j, the agent achieves the set of goals $achieve_i(j) = goals(i) \cap achieve(j)$. Clearly, agent $i \in A$ is most motivated to achieve those sets of goals ranked highest by the ordering $\succeq_i$. However the sets of goals an agent may actually achieve are constrained by both the joint actions in which the agent may participate, J_i, and the function $achieve$. Thus an agent can do no better than to achieve one of the best non-empty set of goals according to $\succeq_i$ given J_i and $achieve$, such sets may be termed *non-dominated*.

Definition 5.2. *For agent $i \in A$ the set of goals $g \in \mathcal{P}(goals(i))$ is non-dominated under $\succeq_i$ if there exists some joint action $j \in J_i$ such that $achieve_i(j) = g$ and there is no $j' \in J_i$ such that $achieve_i(j') \succ_i g$. We denote by max_i, where $max_i \subseteq \mathcal{P}(goals(i)) \setminus \emptyset$, the set of goal sets that are non-dominated under $\succeq_i$ given J_i and $achieve$.*

Agents may have many non-dominated goal sets. Note that our definition of non-domination considers only those goal sets which can actually be achieved through participation in a joint action. Hence for $g, g' \in \mathcal{P}(goals(i))$ where $g \in max_i$ this does not preclude that $g' \succ_i g$. However, since g is non-dominated there exists no $j \in J_i$ such that $achieve_i(j) = g'$. Thus an agent can do no better than to achieve one of its non-dominated goal sets. It follows from our intuitive understanding that for joint actions achieving a non-dominated goal set the agents should exhibit unconditional support. We have:

Definition 5.3. ***Postulate of unconditional support*** *- for agent $i \in A$ and joint action $j \in J_i$, if the set of goals $achieve_i(j) \in max_i$ then $q(i, j) = 0$. Agents unconditionally support joint actions achieving their non-dominated goal-set(s).*

Whilst the postulate of unconditional support stipulates that rational agents must unconditionally support a joint action achieving one of their non-dominated goal-sets the postulate does not limit unconditional support to *only* those joint actions achieving such a goal-set. Even where some set of goals is strictly preferred to some other it may still be rational for an agent to unconditionally support the latter set of goals, for example; where those goals are still 'good enough'. We accommodate such cases as follows:

Definition 5.4. ***Postulate of consistent unconditional support*** *- for agent $i \in A$ and joint actions $j, j' \in J_i$ where $achieve_i(j) \succeq_i achieve_i(j')$ if $q(i, j') = 0$ then $q(i, j) = 0$. If agent i unconditionally supports joint action j' then i also unconditionally supports any joint action that is 'equal to or better than' j'.*

We now consider the other extreme, those joint actions for which the agent has no motivation to participate in. There is no motivation for an agent to participate in a joint action where that joint action achieves none of the agent's goals. It follows that a rational agent will object to participation in a joint action through which it achieves nothing:

Definition 5.5. ***Postulate of objection*** *- for agent $i \in C \subseteq A$ and joint action $j \in J_i$, where j is such that $achieve_i(j) = \emptyset$ then $\frac{|C|-1}{|C|} < q(i, j) \leq 1$. Agents object to joint actions which achieve none of their goals.*

The postulate of objection has an important consequence. Agents must object to participation in any joint action which does not result in them achieving any of their goals. Therefore, where the quorum function, q is rationally specified, it is the case that every consensus action (strong or weak) achieves at least one goal for each participating agent.

As for the postulate of unconditional support, the postulate of objection does not specify where an agent should not object to participation in a joint action. For example, an uncompromising, yet still rational agent may object to any joint action not achieving one of its non-dominated goal sets. We accommodate such cases, thus:

Definition 5.6. ***Postulate of consistent objection*** *- for agent $i \in A$ and joint actions $j, j' \in J_i$ by coalitions $C, C' \subseteq A$ respectively, where $achieve_i(j) \succeq_i achieve_i(j')$ if $\frac{|C|-1}{|C|} < q(i, j) \leq 1$ then $\frac{|C'|-1}{|C'|} < q(i, j') \leq 1$. If agent i objects to j then i also objects to any joint action that is 'no better than' j.*

Our final postulate, that of proportional support, addresses directly the intuition that, a higher quorum value suggests that an agent is less keen to participate in some joint action. Consider the joint actions $j, j' \in J_i$ performed, respectively, by coalitions $C, C' \subseteq A$ and where $q(i, j) = 0.6$, $q(i, j') = 0.7$. From our intuition we may conclude that agent i is more keen to participate in j than in j'; however, this is not necessarily the case. Were it that $|C| = 2$ and $|C'| = 4$, then clearly i is less keen to participate in j that in j' i.e., i objects to j yet conditionally supports j'. Therefore our intuition is only partial. We can now state:

Definition 5.7. ***Postulate of proportional support*** *- for agent $i \in A$ and joint actions $j, j' \in J_i$, by coalitions C and C' respectively, where $|C| \geq |C'|$ and $achieve_i(j) \succeq_i achieve_i(j')$ then $q(i, j) \leq q(i, j')$. Agents' quorum values increase with weak monotonicity as their keenness to participate in joint actions decreases, but only for coalitions of similar or greater cardinality.*

5.3 Computational Complexity of CAG-Gs

Continuing our characterisation of the computational complexity of our models, this section considers the representation and computational complexity of decision problems associated with CAG-Gs. We begin by observing that the addition of goals and agents' orderings over subsets of goals to the CAG model and the subsequent constraints for rational specification of the quorum function has no impact on the complexity of decision problems previously considered in section 4.4. Therefore these results also hold for CAG-Gs.

We therefore focus our attention upon decision problems specifically related to the achievement of goals in CAG-Gs. For CAG-Gs, where agents have an ordering over sets of goals it is natural that we are interested in the ability of agents to achieve one of their non-dominated goal sets. Therefore we will consider the computational complexity of decision problems associated with the verification, existence and non-existence of consensus joint actions in which every participating agent achieves one of their non-dominated goal sets.

We begin, as before setting out the representational scheme to be used as the input for our new decision problems.

5.3.1 Representation

The representational scheme, R, for CAG-Gs extends our representation for CAGs. We must include a representation of the goals achieved by each joint action and the orderings of the agents over subsets of goals. Previous assumptions are carried forward, in particular, our results assume the non-deterministic random access machine (NRAM) model of computation. See section 4.4.1 for further details.

A naive encoding of the agents' ordering over sets of goals would represent each ordering as a set of pairs. Such a representation would include the pair $(x, y) \in 2^G \times 2^G$ for each $(x, y) \in \succeq_i$ and for each $i \in A$; this encoding, although straightforward, has size exponential in the number of goals (space complexity of $O(n \times 2^G \times 2^G)$). We wish to, wherever possible, express complexity results as a function of the number of agents given in the input. The space complexity of this naive representation would obfuscate such results.

Instead, we encode only the relevant part of the agents' ordering over sets of goals in a manner that is more suitable to the decision problems we are going to consider and also leverages the random access model. In particular, for each possible subset of goals we represent the subset of agents for whom those goals are non-dominated. Formally, for the subset of goals $g \in \mathcal{P}(G)$ let $C_g \subseteq A$ be a set of agents such that agent $i \in C_g$ only if g is non-dominated under $\succeq_i$; then the set of agents for whom g is non-dominated is represented as $rep(g) = \{i \in C_g \subseteq A \mid (g \cap goals(i)) \in max_i\}$. Let $rep(G) = \{rep(g) \mid g \in \mathcal{P}(G)\}$ be an ordered sequence of sets of agents, indexed by the set of goals $g \in \mathcal{P}(G)$. We assume that the structure $rep(G)$ is indexed such that given a set of goals $g \in \mathcal{P}(G)$ the set of agents $C_g \subseteq A$, for which g is a set of non-dominated goals can be located in constant time. This representation has been chosen so that we may present meaningful complexity results for the decision problems we are going to consider.

We extend our encoding of joint actions to include both the quorum values of the agents for each joint action and the goals achieved through each joint action. As before, let $Ag : J \longrightarrow \mathcal{P}(A)$ be a function that for a given joint action j returns $C \subseteq A$, the set of agents that may participate in joint action j. The joint action j is fully described by the set of quadruples $rep(j) = \{(i, j, q(i, j), achieve(j)) \mid i \in Ag(j), q(i, j) \in [0, 1], achieve(j) \in \mathcal{P}(G)\}$. The overall representation of the game is then the ordered sequence given by the sets $R = \{\{rep(j) \mid j \in J)\}, \{rep(G)\}\}$.

5.3.2 Complexity of strong consensus decision problems

STRONG CONSENSUS NON-DOMINATED (SCN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ and joint action $j \in J$, is j a strong consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

The algorithm must verify that j achieves a non-dominated goal-set for each agent $i \in Ag(j)$. This can be achieved by looking up the set of agents C_x indexed by $x = achieve(j)$ in the indexed structure $rep(G)$ and verifying that $Ag(j) \subseteq C_x$. This is an $O(n)$ operation. If j achieves a non-dominated goal set for each agent then, under the postulate of unconditional support j is a strong consensus action (since $\forall i \in Ag(j), q(i, j) = 0$). This gives a deterministic algorithm with time complexity $O(n)$. Hence the problem is in $P(n)$ for NRAM.

STRONG CONSENSUS NON-DOMINATED EXISTS (SNE)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ does there exist some $j \in J$ such that j is a strong consensus action and each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, worst case, evaluate SCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$

A non-deterministic algorithm first guesses an index of an action j in J, this is an $O(log(|J|)$ operation. The algorithm then checks that through j each participating agent achieves one of their non-dominated goal-sets, this is an $O(n)$ operation. This gives a non-deterministic linear time algorithm for a random access machine. Hence the problem is in $NP(n)$ for NRAM.

NO STRONG CONSENSUS NON-DOMINATED (SNN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ is there no $j \in J$ such that j is a strong consensus action and each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, worst case, evaluate the complement of SCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$

Since the complement of this problem (SNE) is in $NP(n)$, the non-deterministic algorithm for this problem is in *co*-$NP(n)$

STRONG CONSENSUS GROUP NON-DOMINATED (SGN)

Given a CAG-G $G = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$, and a coalition $C \subseteq A$, does there exist some $j \in J_C$ such that j is a strong consensus action and each agent $i \in C$ achieves one of their non-dominated goal-sets?

Since it is possible that $\forall j \in J, Ag(j) = C$ this problem has identical complexity to SNE. The deterministic algorithm therefore has time complexity $O(n \times |J|)$

Since this problem is similar to SNE a non-deterministic algorithm for this problem is also in $NP(n)$ for NRAM.

q_C-MINIMAL STRONG CONSENSUS NON-DOMINATED (QM-SCN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$, and a joint action $j \in J$, is j a q_C-minimal strong consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

Recall that:

- the q_C-minimal core is defined as containing only those joint actions for which the number of agents and the number of rounds required for strong consensus are minimal. Specifically, a strong consensus action j by the agents $C \subseteq A$ is a q_C-*minimal consensus action* if there is no other strong consensus action $j' \neq j$ for $C' \subseteq C$ where the number of rounds required to reach consensus for j' is less than the number of rounds required to reach consensus regarding j.

- $rounds : J \rightarrow \mathbb{N}$ is a function giving the number of rounds required for the agents to reach strong consensus regarding joint action $j \in J$ and has time complexity $O(n)$.

- $J_{C' \subseteq C} = \{j \in J \mid \{i \mid (i, a) \in j\} \subseteq C\}$ denotes the set of all joint actions for all groups of agents $C' \subseteq C$.

A deterministic algorithm must verify that for all agents $i \in Ag(j)$, $achieve(j)$ is a non-dominated goal set (this is an $O(n)$ operation). If j achieves a non-dominated goal set for each participating agent then, under the postulate of unconditional support, j is a strong consensus action and consensus is reached in one round. Since the minimum number of rounds to reach consensus is one round then j is also in the q_C-minimal core (there can be no $j' \neq j$ for which strong consensus can be reached in strictly fewer rounds). A deterministic algorithm therefore has time complexity $O(n)$. Hence the problem is in $P(n)$.

q_C-MINIMAL STRONG CONSENSUS NON-DOMINATED EXISTS (QM-SNE)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ is there some $j \in J$ such that j is a q_C-minimal strong consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, in the worst case evaluate QM-SCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first guesses an index of some j in J, the procedure then follows as for QM-SCN. Therefore this problem is also in $NP(n)$ for NRAM.

NO q_C-MINIMAL STRONG CONSENSUS NON-DOMINATED (QM-SNN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ is there no $j \in J$ such that j is a q_C-minimal strong consensus action and that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, in the worst case evaluate the complement of QM-SCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$.

Since the complement of this problem (QM-SNE) is in $NP(n)$, this problem is in $co\text{-}NP(n)$ for NRAM.

q_C-MINIMAL STRONG CONSENSUS GROUP NON-DOMINATED (QM-SGN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ and a subset of agents $C \subseteq A$ is there some $j \in J_C$ such that j is a q_C-minimal strong consensus action and each agent $i \in C$ achieves one of their non-dominated goal-sets?

A deterministic algorithm must verify that $\exists j \in J_C$ such that j is a strong consensus action through which each participating agent achieves one of their non-dominated goal-sets and that $\neg\exists j' \in J_{C' \subseteq C}, j \neq j'$ such that j' is also a strong consensus action where consensus is reached in strictly less rounds for j' than for j. Therefore the problem has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first calls an $NP(n)$ oracle to check that C has a strong consensus action, j and that $\forall i \in C$, $achieve(j)$ is non-dominated; if there is such an action for C it then calls an $NP(n)$ oracle to check whether any $C' \subseteq C$ has a strong consensus action converging in strictly fewer rounds. Hence the problem is in $D^p(n)$ for NRAM.

5.3.3 Complexity of weak consensus decision problems

WEAK CONSENSUS NON-DOMINATED (WCN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$, and a joint action $j \in J$, is j a weak consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm must verify that j achieves a non-dominated goal-set for each agent $i \in Ag(j)$; this can be achieved by looking up $x = achieve(j)$ in the indexed structure $rep(G)$ and verifying that $Ag(j) \subseteq C_x$. This is an $O(n)$ operation. From the postulate of unconditional support, if j achieves a non-dominated set of goals for each participating agents then $\forall i \in Ag(j), q(i, j) = 0$, hence j is a weak consensus action. This gives a deterministic algorithm with time complexity $O(n)$. Hence the problem is in $P(n)$ for NRAM.

WEAK CONSENSUS NON-DOMINATED EXISTS (WNE)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i \mid i \in A\}, q\rangle$ is there some $j \in J$ that is a weak consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, worst case, evaluate WCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$.

A non-deterministic algorithm first guesses an index of an action j in J, this is an $O(log(|J|)$ operation. The algorithm then checks that through j each participating agent achieves one of their non-dominated goal-sets, this is an $O(n)$ operation, if so then j is a weak consensus action. This gives an non-deterministic linear time algorithm for a random access machine. Hence the problem is in $NP(n)$ for NRAM.

NO WEAK CONSENSUS NON-DOMINATED (WNN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ is it the case that there is no $j \in J$ such that j is a weak consensus action and that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, worst case, evaluate the complement of WCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|)$.

Since the complement of this problem (WNE) is in $NP(n)$, the non-deterministic algorithm for this problem is in *co*-$NP(n)$.

WEAK CONSENSUS GROUP NON-DOMINATED (WGN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ and a subset of agents $C \subset A$ is there a weak consensus action for C such that each agent $i \in C$ achieves one of their non-dominated goal-sets?

Since it is possible that $\forall j \in J, Ag(j) = C$ this problem has identical complexity to WNE. The deterministic algorithm therefore has time complexity $O(n \times |J|)$.

Since this problem is similar to WNE a non-deterministic algorithm for this problem is also in $NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS NON-DOMINATED (CM-WCN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ and a joint action $j \in J$ is j a C-minimal consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

Recall that: a weak consensus action $j \in J_C$ is a *C-minimal consensus action* and hence in the C-minimal core of a CAG only if there does not exist some other weak consensus action $j' \in J_{C'}$ such that $C' \subset C$.

A deterministic algorithm must verify that $achieve(j)$ gives a non-dominated goal set for each $i \in Ag(j)$ (this is an $O(n)$ operation). If so the algorithm must then verify that $\neg\exists j' \in J_{C' \subset C}, j \neq j'$ such that j' is also a weak consensus action (this is an $O(n \times |J|)$ operation). Hence the problem has time complexity $O(n \times |J|)$.

A non-deterministic algorithm for deciding the *complement* of this problem will first check whether through j each agent achieves one of their non-dominated goal sets (and return yes if they do not) and if it is, guess a joint action $j' \in J$, verify that $j' \neq j$, $Ag(j') \subset Ag(j)$ and j' is a weak consensus action. The problem of deciding that j is *not* a C-minimal consensus action through which each participating agent achieves one of their non-dominated goal-sets is therefore in $NP(n)$ on NRAM. Hence deciding whether j is a C-minimal consensus action through which each participating agent achieves one of their non-dominated goal-sets is in *co*-$NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS NON-DOMINATED EXISTS (CM-WNE)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ is there some $j \in J$ that is a C-minimal consensus action such that each agent $i \in Ag(j)$ achieves one of their non-dominated goal-sets?

A deterministic algorithm would, in the worst case evaluate CM-WCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|^2)$.

A non-deterministic algorithm first guesses an index of some j in J, after-which the procedure is as for CM-WCN. Therefore this problem is also in *co*-$NP(n)$ for NRAM.

NO C-MINIMAL WEAK CONSENSUS NON-DOMINATED (CM-WNN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ is it the case that no $j \in J$ is a C-minimal consensus action through which each participating agent achieves one of their non-dominated goal-sets?

A deterministic algorithm would, in the worst case evaluate the complement of CM-WCN for each $j \in J$. The algorithm therefore has time complexity $O(n \times |J|^2)$.

Since the complement of this problem (CM-WNE) is in *co*-$NP(n)$, this problem is in $NP(n)$ for NRAM.

C-MINIMAL WEAK CONSENSUS GROUP NON-DOMINATED (CM-WGN)

Given a CAG-G $\Gamma = \langle A, Ac, J, G, goals, achieve, \{\succeq_i | \ i \in A\}, q\rangle$ and a subset of agents $C \subseteq A$, is there a C-minimal consensus action for C such that each agent $i \in C$

achieves one of their non-dominated goal-sets?

Since it is possible that $\forall j \in J, Ag(j) = C$ this problem has identical complexity to GM-WNE. The deterministic algorithm therefore has time complexity $O(n \times |J|^2)$.

A non-deterministic algorithm first calls an $NP(n)$ oracle to check that C has a weak consensus action through which each participating agent achieves one of their non-dominated goal-sets. If so, it then calls an $NP(n)$ oracle to check whether any $C' \subset C$ has a weak consensus action. Hence the problem is in $D^p(n)$ for NRAM.

5.4 Comparison of CAGs and CAG-Gs

The CAG-G model extends that of CAGs by the introduction of a set of goals, G, a function mapping agents $i \in A$ to subsets of goals, $goals(i) \subseteq G$, a function mapping joint actions $j \in J$ to the goals they achieve, $achieve(j) \subseteq G$ and, for each agent $i \in A$ the pre-order $\succeq_i \subseteq 2^{goals(i)} \times 2^{goals(i)}$. The introduction of these structures has no effect on the solution concepts previously defined for CAGs. Hence, the solution concepts of strong/weak consensus actions and the associated concepts of the q_C-minimal core and the C-minimal core defined for CAGs apply to CAG-Gs without alteration.

Although agents in CAG-Gs possess an ordering over subsets of their goals in CAG-Gs we do not attempt to aggregate these orderings into an overarching, total social order as in social choice theory. Instead, these orderings place rational constraints upon the specification of the quorum function, q. Like CAGs, consensus and core membership in CAG-Gs are evaluated using only the thresholds given by q.

Our complexity results, summarised in table 5.1 show that the impact of extending the CAG model to CAG-Gs is small, yet noticeable. Of the sixteen decision problems, twelve of these exhibited similar time complexity and membership of similar complexity classes to those found for related problems in CAGs. The comparability of these results is due largely to our choice of representation and indexing for $rep(G)$. Since establishing that a subset of goals are non-dominated for some subset of agents is an $O(n)$ operation we might expect there to be little impact on the complexity of decision problems.

However, for at least one decision problem there is a reduction in complexity. In CAGs the decision problem QM-SCA asks whether a given joint action $j \in J$ is in the q_C-minimal core, this problem was found to be in $NP(n)$. The comparable decision problem in CAG-Gs is QM-SCN, determining whether j achieves a non-dominated goal set for each participating agent and is also q_C-minimal was found to be in $P(n)$.

The apparent simplification of this problem is due to the introduction of rationality postulates in CAG-Gs. Specifically the postulate of unconditional support guarantees that if j achieves a non-dominated set of goals for each participating agent then $\forall i \in Ag(j), q(i,j) = 0$ and $rounds(j) = 1$. Therefore, there cannot exist some other $j' \in J$ such that $rounds(j') < rounds(j)$. Hence j must be in the q_C-minimal core for CAG-Gs.

The complexity of decision problems CM-WEA, CM-WNA and CM-WGA is also different in CAG-Gs to the comparable problems in CAGs. These problems are concerned with the existence and non-existence of joint actions in the C-minimal core, their time complexity has increased from $O(n \times |J|)$ to $O(n \times |J|^2)$. These differences can be directly associated with the introduction of goals in the CAG-G model. In CAGs the problems of existence/emptiness of the C-minimal core can be reduced to the problems of (non)-existence of any weak consensus action. Such a reduction is not possible for CAG-Gs since the existence of a weak consensus action, $j \in J$ does not guarantee that j achieves a non-dominated goal set for each participating agent.

Table 5.1: Comparison of complexity results for CAGs and CAG-Gs.

Consensus Action Games			**Consensus Action Games with Goals**		
Decision Problem	Time Complexity	Complexity Class	Decision Problem	Time Complexity	Complexity Class
SCA	$O(n)$	$P(n)$	SCN	$O(n)$	$P(n)$
SAE	$O(n \times |J|)$	$NP(n)$	SNE	$O(n \times |J|)$	$NP(n)$
SNA	$O(n \times |J|)$	$co\text{-}NP(n)$	SNN	$O(n \times |J|)$	$co\text{-}NP(n)$
SGA	$O(n \times |J|)$	$NP(n)$	SGN	$O(n \times |J|)$	$NP(n)$
QM-SCA	$O(n \times |J|)$	$co\text{-}NP(n)$	QM-SCN	$O(n)$	$P(n)$
QM-SAE	$O(n \times |J|)$	$NP(n)$	QM-SNE	$O(n \times |J|)$	$NP(n)$
QM-SNA	$O(n \times |J|)$	$co\text{-}NP(n)$	QM-SNN	$O(n \times |J|)$	$co\text{-}NP(n)$
QM-SGA	$O(n \times |J|)$	$D^p(n)$	QM-SGN	$O(n \times |J|)$	$D^p(n)$
WCA	$O(n)$	$P(n)$	WCN	$O(n)$	$P(n)$
WAE	$O(n \times |J|)$	$NP(n)$	WNE	$O(n \times |J|)$	$NP(n)$
WNA	$O(n \times |J|)$	$co\text{-}NP(n)$	WNN	$O(n \times |J|)$	$co\text{-}NP(n)$
WGA	$O(n \times |J|)$	$NP(n)$	WGN	$O(n \times |J|)$	$NP(n)$
CM-WCA	$O(n \times |J|)$	$co\text{-}NP(n)$	CM-WCN	$O(n \times |J|)$	$co\text{-}NP(n)$
CM-WAE	$O(n \times |J|)$	$NP(n)$	CM-WNE	$O(n \times |J|^2)$	$co\text{-}NP(n)$
CM-WNA	$O(n \times |J|)$	$co\text{-}NP(n)$	CM-WNN	$O(n \times |J|^2)$	$NP(n)$
CM-WGA	$O(n \times |J|)$	$D^p(n)$	CM-WGN	$O(n \times |J|^2)$	$D^p(n)$

5.5 Discussion and Related Work

In this chapter we have described consensus action games with goals (CAG-Gs). CAG-Gs are coalitional games in which self-interested agents achieve individual goals through the formation of coalitions and performance of joint actions. The CAG-G model describes, for each agent $i \in A$, an ordering, $\succeq_i$, over subsets of the agent's goals. Thus, for each agent, the achievement of certain combinations of goals may be more desirable than others. Given such an arrangement the traditional game-theoretic treatment would be to derive expressions of coalitional stability directly from the agents' preference orderings. This is the approach taken by Dunne and Wooldridge when formulating qualitative coalition games with preferences (QCGPs) [Dunne and Wooldridge, 2004].

Recall (from section 3.9) that in a qualitative coalition game (QCG) agents form coalitions in order to achieve their individual goals; it is assumed that agents are indifferent between which of their goals are achieved. In QCGPs this assumption of indifference is dropped. Each agent holds a preference ordering over their goals, from these orderings Dunne and Wooldridge define notions of strong and weak coalitional preference. A coalition strongly prefers some set of goals where that set contains at least one goal, common to all members; for weak coalitional preference the requirement that the most preferable goal to be common to all member-agents is relaxed.

CAG-Gs differ from QCGPs in several respects. Firstly, in CAG-Gs we include a representation of actions and joint actions for the agents; since the focus of QCGPs is solely on goal achievement, QCGPs have nothing to say about what the agents will actually do. In QCGPs agents' preferences are over their set of goals, thus there is an implicit assumption in QCGPs that goal preference is weakly monotonic in the number of goals each agent achieves (that is: achieving an additional goal never leads to a less preferred outcome). Since in CAG-Gs agents' orderings are over subsets of their goals we do not make this assumption. The cost of this increased expressivity is that, potentially, the structure required to represent agents' ordering over goals in CAG-Gs is exponential in the size of an agent's goal set, whereas, for QCGPs the comparable structure is quadratic in the size of the same set. However, since we do not require that $\succeq_i$ is total it is not necessarily the case that for similar goal-sets the space complexity of preference representation in CAG-Gs will always be greater than for QCGPs.

Most strikingly however, in CAG-Gs we do not follow the expected game theoretic route of deriving coalition stability directly from preference. Instead the agents' goal-set orderings place a series of linear constraints on the specification of the quorum function, q. These constraints are formulated in accordance with the five rationality postulates given in section 5.2, these postulates concern themselves with the rational selection of threshold values for the agents.

Since rationality is central to the study of games, our treatment of agents' preference in CAG-Gs is not as divergent from the typical approach as it may first appear. In Harsanyi [1961] a series of six postulates concerning rationality in two person co-operative games with transferable utility are proposed. Although Harsanyi's principal concern is with decision making in the bargaining or negotiation problem it turns out that our approach to rationality is largely consonant with rationality in this economic setting. Harsanyi's postulates may be summarised as follows:

- individual utility maximisation - each player tries to get the best possible outcome for themselves,
- acceptance of higher pay-offs - if a player makes an offer in which they receive a pay-off of some amount, that player will settle for any counter-offer where they receive at least that amount,
- efficiency - given a set of decision rules players will choose those rules that are most efficient in the sense that they lead to higher pay-offs for both players,
- symmetry - the same decision rules apply to all players,
- restriction of variables - decision rules need not be concerned with anything other than the coalitional reward and its subsequent division,
- mutually expected rationality - players expect that other players will follow decision rules consistent with these postulates.

The postulate of individual utility maximisation characterises players as principally self-interested. In CAG-Gs self interest is served by unconditionally supporting those joint actions which achieve the most preferable goal-sets. Hence, Harsanyi's postulate of individual utility maximisation is directly analogous to our postulate of unconditional support (definition 5.3).

The postulate of acceptance of higher pay-offs characterises players as having consistent behaviour and is analogous to our postulates of proportional support (definition 5.7) and consistent unconditional support (definition 5.4). The implicit converse postulate (non-acceptance of lower pay-offs) is therefore analogous to our postulate of consistent objection (definition 5.6). It is also the case that CAG-G's satisfy Harsanyi's postulates of symmetry and mutually expected rationality; in CAG-Gs the same rules and definitions of rationality apply to all players.

Since CAG-G's do not attempt to represent situations where utility is transferable (i.e., goals are personal thus one agent cannot achieve the goals of another) it is less straightforward to find parallels between rationality in CAG-G's and Harsanyi's postulates of efficiency and restriction of variables - both are concerned in some-way with coalitional rewards. Nonetheless, CAGs are efficient in the sense that where agents achieve their non-dominated goal sets ("highest pay-off") the agents will unconditionally support this outcome. It is also the case that our rationality postulates for CAG-Gs are concerned only with the goals that may be achieved, thus, in this sense we observe Harsanyi's postulate of restriction of variables.

Chapter 6

Temporal Consensus Action Games with Goals

6.1 Temporal Consensus Action Games with Goals

In this chapter we present our final model, temporal consensus action games with goals (T-CAG-Gs). The T-CAG-G model extends previous models by adding a temporal element to the games. T-CAG-Gs describe agents playing, possibly different CAG-Gs repeatedly, over time.

A T-CAG-G is a transition system over a set of states, T. Each state $t \in T$ is labelled with:

- a CAG-G game, $\gamma(t)$, over common sets of agents, A, actions, Ac, joint actions, J, and goals, G,
- for each goal in G a goal proposition, with valuation function $\pi_G(t)$,
- a set of atomic propositions, P, with valuation function $\pi_P(t)$.

A goal proposition is true in a given state only if that goal is achieved in that state. The set of propositions, P, is arbitrary and may be used to model the domain of interest.

Our intention is that T-CAG-Gs be suitable for modelling real-world problems. As such we expect that in many applications for T-CAG-Gs the domain of interest will obey certain domain specific constraints. In T-CAG-Gs these constraints are captured using the structure C, a set of propositional formulae. Constraints may be placed on which goal and other propositions may be concurrently true. This may be useful, for example, when modelling the location of a physical object; propositions representing the object's location are mutually exclusive since an object may not occupy two places concurrently.

In T-CAG-Gs, as with previous models, a game may possess multiple consensus actions. The transition relation, o, for T-CAG-Gs respects this, permitting multiple joint actions to be performed simultaneously, thus affecting the choice of subsequent state. The T-CAG-G model therefore describes circumstances where two or more coalitions of agents act with 'internal' consensus — there is consensus amongst the agents of each coalition, but without 'external' consensus – there is no consensus between those coalitions. For these cases we require that: the participating coalitions are disjoint with respect to agents (an agent cannot be a member of two coalitions in T-CAG-Gs) and that their combined actions do not violate those domain constraints given in C. State transitions in T-CAG-Gs therefore represent one or more joint actions performed by disjoint coalitions. We refer to the concurrent performance of multiple joint actions as *composite actions*. More formally:

Definition 6.1. *A T-CAG-G over a given set of agents A, goals G, and joint actions J is a structure $\langle A, Ac, J, G, achieve, P, C, T, \gamma\pi_G, \pi_P, \alpha, o\rangle$ where:*

A *is a finite set of agents,* $\{1, \ldots, n\}, n \geq 2$.

Ac *is a finite non-empty set of possible actions* $\{1, \ldots, m\}$.

J *is a set of joint actions; each joint action is a set of pairs (i, a), where $i \in A$ and $a \in Ac$, specifying the action performed by each agent participating in the joint action. Each participating agent performs exactly one action.*

G *is a finite non-empty set of goal propositions.*

$achieve$ *is a function $achieve : J \longrightarrow 2^G$. For each joint action $j \in J$, $achieve(j) \subseteq G$ gives the set of goals achieved through the performance of j.*

P *is a finite, possibly empty, set of atomic propositions.*

C *is a set of constraints, each defined as a propositional formula over the set of propositions $G \cup P$.*

T *is a finite non-empty set of states. Each state is labelled with a CAG-G and the set of propositions $G \cup P$.*

Let S be a set of all possible CAG-Gs of the form $\langle A, Ac, J_t, G, goals_t, achieve, \{\succeq_{i,t} \mid i \in A\}, q_t\rangle$ where $t \in T$ and $J_t \subseteq J$. Components which are indexed by $t \in T$ are allowed to vary while non-indexed components are common for all elements of S.

γ *is a function $\gamma : T \longrightarrow S$ assigning to each state $t \in T$ a CAG-G $\gamma(t) \in S$.*

π_G *is a function $\pi_G : T \longrightarrow 2^G$ assigning to each state $t \in T$ a set of goal propositions which are true (achieved) in t.*

π_P *is a function $\pi_P : T \longrightarrow 2^P$ assigning to each state $t \in T$ a set of propositions from P which are true in t.*

$\alpha(t)$ *is a function $\alpha : T \longrightarrow 2^J$ assigning to the CAG-G $\gamma(t)$ a set of joint actions, $\alpha(t) = J_t$, the joint actions possible in t.*

Let E_t be a set of sets of joint actions applicable in state $t \in T$. Each $c \in E_t$ is a composite action $c = \{j_1, \ldots, j_k\}$, $k \leq n$, $j_1, ..., j_k \in \alpha(t)$. We say c is applicable only if:

$Ag(j_a) \cap Ag(j_b) = \emptyset$ for all $a, b \in 1...k, a \neq b$, ($Ag : J \longrightarrow 2^A$) and

for any $g, g' \in achieve(j_1) \cup ... \cup achieve(j_k), g \neq g'$ the expression $\neg(g \wedge g')$ cannot be derived from the constraints given in C.

We require that $\forall c \in E_t$, c is applicable. We denote the set of all composite actions as $E = \bigcup_{t \in T} E_t$.

o *is a serial transition relation. For state $t \in T$ and applicable composite action $c \in E_t$, $o(t, c) \in T$ is the successor state reached by performing the composite action c in state t. Since o is serial we require that $\forall t \in T, o(t, \emptyset) = t$. The relation o, is also constrained such that for $c = \{j_1, \ldots, j_k\}$, $\pi_G(o(t, c)) \supseteq achieve(j_1) \cup ... \cup achieve(j_k)$. At least those goals achieved by each joint action $j \in c$ are achieved (true) in the successor state.*

6.2 Temporal Solution Concepts for T-CAG-Gs

In this section we define temporal solution concepts for T-CAG-Gs. Recall that T-CAG-Gs are a transition system in which each transition corresponds to a composite action. For T-CAG-Gs a solution is a full (infinite or terminated) path through this transition system. A solution concept for T-CAG-Gs is a solution where each transition is a composite action composed only of consensus joint actions. We refer to a composite action c as a *composite consensus action* only if $\forall j \in c$, j is stable under a given temporal solution concept.

Our temporal solution concepts are derived from individual rationality in consensus games (Chapter 3) and collective rationality in consensus action games (Chapter 4). We refer to our temporal solution concepts for T-CAG-Gs as *consensus paths* and define these below.

Definition 6.2. *(Strong consensus path) — A path is a strong consensus path if and only if it is a full path and each transition of the path corresponds to a composite action composed entirely of strong consensus actions. That is, each transition, $(t, t') \in T \times T$, of the path corresponds to a $c \in E_t$ such that $\forall j \in c$, j is a strong consensus action.*

Definition 6.3. *(Weak consensus path) — A path is a weak consensus path if and only if it is a full path and each transition, $(t, t') \in T \times T$, of the path corresponds to a $c \in E_t$ such that $\forall j \in c$, j is a weak consensus action.*

Definition 6.4. *(q_C-Strong consensus path) — A path is a q_C-strong consensus path if and only if it is a full path and each transition, $(t, t') \in T \times T$, of the path corresponds to a $c \in E_t$ such that $\forall j \in c$, j is a q_C-minimal strong consensus action.*

Definition 6.5. *(C-Weak consensus path) — A path is a C-weak consensus path if and only if it is a full path and each transition, $(t, t') \in T \times T$, of the path corresponds to a $c \in E_t$ such that $\forall j \in c$, j is a C-minimal weak consensus action.*

6.3 Example T-CAG-G

Example 6.1. *Alice (Al) and Bob (Bo) marry and decide to move into a home of their own. We find them in the hallway of their new home as they are deciding where to put a box of books. The box is quite heavy, and although either of them could lift the box it would be easier if they lift the box together. Neither wants the box to remain in the hallway; Bob would like to move the the box into the study (s), whilst Alice would like the box to be in the library (l). We formalise this example as the temporal consensus action game $\Gamma_4 = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o \rangle$ where:*

$A = \{Al, Bo\}$

$Ac = \{s, l\}$

$J = \{j_1, \ldots, j_6\}$ *where:*

$j_1 = \{Al^s\}$ *— Alice moves the box to the study,*

$j_2 = \{Al^l\}$ *— Alice moves the box to the library,*

$j_3 = \{Bo^s\}$ *— Bob moves the box to the study,*

$j_4 = \{Bo^l\}$ *— Bob moves the box to the library,*

$j_5 = \{Al^s, Bo^s\}$ *— they both move the box to the study,*

$j_6 = \{Al^l, Bo^l\}$ *— they both move the box to the library.*

$G = \{g_h, g_s, g_l\}$ *where:*

g_h *is the goal that the box is in the hallway,*

g_s *is the goal that the box is in the study,*

g_l *is the goal that the box is in the library.*

$achieve : J \longrightarrow 2^G$ *where:*

$achieve(j_1), achieve(j_3), achieve(j_5) = \{g_s\}$,

$achieve(j_2), achieve(j_4), achieve(j_6) = \{g_l\}$.

$P = \emptyset$

$C = \{\phi_1, \ldots, \phi_3\}$

ϕ_1 *is* $g_h \iff \neg(g_s \vee g_l)$,

ϕ_2 *is* $g_s \iff \neg(g_h \vee g_l)$,

ϕ_3 *is* $g_l \iff \neg(g_h \vee g_s)$.

$T = \{t_0, t_1, t_2\}$ *There are three states, in each exactly one* $g \in G$ *is true. Associated with state* t_0 *is the CAG-G* $\gamma(t_0)$*, where:*

$goals_{t_0}(Al) = \{g_l\}$ *and* $goals_{t_0}(Bo) = \{g_s\}$

$\succeq_{Al}$ *ranks* $\{g_l\} > \{g_s\} > \{g_h\} > \emptyset$ *and* $\succeq_{Bo}$ *ranks* $\{g_s\} > \{g_l\} > \{g_h\} > \emptyset$

$$q_{t0}(i,j) = \begin{cases} 1 & \textit{if } i = Al \textit{ and } j = j_1 \\ 0 & \textit{if } i = Al \textit{ and } j = j_2 \\ 0 & \textit{if } i = Bo \textit{ and } j = j_3 \\ 1 & \textit{if } i = Bo \textit{ and } j = j_4 \\ 1 & \textit{if } i = Al \textit{ and } j = j_5 \\ 0 & \textit{if } i = Bo \textit{ and } j = j_5 \\ 0 & \textit{if } i = Al \textit{ and } j = j_6 \\ 1 & \textit{if } i = Bo \textit{ and } j = j_6 \end{cases}$$

$\pi_G : T \longrightarrow 2^G$ *are the goals true (achieved) in* t

$$\pi_G(t_0) = \{g_h\},$$
$$\pi_G(t_1) = \{g_s\},$$
$$\pi_G(t_2) = \{g_l\}.$$

$\pi_P : T \longrightarrow 2^P$, $\pi_P(t) = \emptyset$ *for all* $t \in T$

$\alpha(t_0) = J$*, for all other states* $\alpha(t) = \emptyset$

$E_{t_0} = \{\{j_1\}, \{j_2\}, \{j_3\}, \{j_4\}, \{j_5\}, \{j_6\}, \{j_1, j_3\}, \{j_2, j_4\}\}$

$$o(t,c) = \begin{cases} t_1 & \textit{if } t = t_0 \textit{ and } c = \{j_1\}, \{j_3\}, \{j_1, j_3\}, \{j_5\} \\ t_2 & \textit{if } t = t_0 \textit{ and } c = \{j_2\}, \{j_4\}, \{j_2, j_4\}, \{j_6\} \\ t & \textit{if } c = \emptyset \end{cases}$$

We begin in state t_0 (the box is in the hallway); in $\gamma(t_0)$ the strong consensus joint actions are: j_2 (Alice moves the box to the library) and j_3 (Bob moves the box to the study). Although these actions are performed by disjoint coalitions ($\{Al\} \cap \{Bo\} = \emptyset$) it would not be possible for both of these equilibria to occur. This is captured through the constraint structure C that encodes domain specific rules; for example: if the box is in the study then the box is not in the hallway or in the library ($g_s \iff \neg(g_h \vee g_l)$). The paths t_0, t_1 and t_0, t_2 are strong consensus paths, in the former the box is moved to the library by Alice, in the latter the box is moved to the study by Bob.

6.4 Computational Complexity of T-CAG-Gs

Our characterisation of the computational complexity of T-CAG-Gs considers the existence of consensus paths from some initial state. This immediately suggests the following decision problems:

STRONG CONSENSUS PATH EXISTS (SPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a strong consensus path from t?

QC-STRONG CONSENSUS PATH EXISTS (QM-SPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a q_C-strong consensus path from t?

WEAK CONSENSUS PATH EXISTS (WPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a weak consensus path from t?

C-WEAK CONSENSUS PATH EXISTS (CM-WPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a C-weak consensus path from t?

6.4.1 Preliminaries

For there to be a consensus path from state t it is necessary (but not sufficient) that t has at least one *consensus successor* state. A state $t \in T$ has a consensus successor only if there exists some joint action $j \in \alpha(t) \subseteq J$ such that, for a given solution concept, j is a consensus action. At this point it will be useful to recall some results from previous chapters. Table 6.1 shows, for each non-temporal solution concept, the complexity of deciding whether $j \in J$ is a consensus action.

Table 6.1: Time complexity of verification problems for non-temporal solution concepts.

Problem	Description	Complexity
SCA	Is j a strong consensus action?	$O(n)$
QM-SCA	Is j a q_C-minimal strong consensus action?	$O(n \times \|J\|)$
WCA	Is j a weak consensus action?	$O(n)$
CM-WCA	is j a C-minimal weak consensus action?	$O(n \times \|J\|)$

In T-CAG-Gs agents may perform composite joint actions. Table 6.2 shows the time complexity of deciding whether the composite action c is a composite consensus action. The input to these decision problems is a T-CAG-G Γ, a state $t \in T$ and a composite action $c \in E_t$. Note that, since composite actions are performed by disjoint sets of agents each composite action is performed by at most n agents. Therefore, these results do not differ from those given for joint actions in table 6.1.

Table 6.2: Time complexity of deciding whether c is a composite consensus action.

Problem	Description	Complexity
SCA-C	Is c a strong composite consensus action?	$O(n)$
QM-SCA-C	Is c a q_C-minimal strong composite consensus action?	$O(n \times \|J\|)$
WCA-C	Is c a weak composite consensus action?	$O(n)$
CM-WCA-C	is c a C-minimal weak composite consensus action?	$O(n \times \|J\|)$

A consensus path consists only of transitions corresponding to composite consensus actions. The problem of determining the existence of a full path along which some invariant property holds has been studied by those interested in the model checking problem. Algorithm 6.1, adapts the algorithm for computing paths with an invariant property [Baier and Katoen, 2008, p. 348] when model checking computation tree logic (CTL) [Emerson and Clarke, 1982]. When the algorithm initialises, the set CP contains all states in T. Upon termination, CP contains only those states on a consensus path.

Algorithm 6.1 terminates in at most $|T|$ iterations, in this case $CP = \emptyset$ and no consensus path exists. In each iteration we check the expression: $\exists t \in CP$ such that t has no consensus successors in CP. To check this expression we must, in the worst case examine every composite action (transition). Recall that $E = \bigcup_{t \in T} E_t$, hence

Algorithm 6.1 Existence of paths with an invariant property.

```
CP ← T
while ∃t ∈ CP such that t has no consensus successors in CP do
    CP ← CP \ {t}
end while
```

the number of transitions in Γ is given by $|E|$ and this check requires time $O(|E|)$. Therefore the overall time required for algorithm 6.1 is $O(|T| \times |E|)$. The algorithm has time complexity that is polynomial in the number of states and number of transitions.

We can now give our complexity results for the four temporal solution concepts.

6.4.2 Complexity of strong decision problems for T-CAG-Gs

STRONG CONSENSUS PATH EXISTS (SPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o \rangle$ and a state, $t \in T$, does there exist a strong consensus path from t?

A deterministic algorithm would iterate through each composite action $c \in E$ labelling those transitions which correspond to strong composite consensus actions. This is an $O(n \times |E|)$ operation. After this pre-computation we apply algorithm 6.1 ($O(|T| \times |E|)$) to compute the set of states that are on consensus paths. Finally we verify that the given state t is found in CP ($O(|T|)$).

The overall time complexity of the deterministic algorithm is $O((n \times |E|) + (|T| \times |E|))$. The algorithm is polynomial in the number of agents, states, and transitions.

q_C-STRONG CONSENSUS PATH EXISTS (QM-SPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o \rangle$ and a state, $t \in T$, does there exist a q_C-strong consensus path from t?

A deterministic algorithm would iterate through each composite action $c \in E$ labelling those transitions which correspond to q_C-minimal strong composite consensus actions. This is an $O((n \times |J|) \times |E|)$ operation (from QM-SCA-C). After this pre-computation we apply algorithm 6.1 ($O(|T| \times |E|)$) to compute the set of states that are on consensus paths. Finally we verify that the given state t is found in CP (an $O(|T|)$ operation).

The overall time complexity of the deterministic algorithm is $O(((n \times |J|) \times |E|) + (|T| \times |E|))$. The algorithm is polynomial in the number of agents, states, joint actions and transitions.

6.4.3 Complexity of weak decision problems for T-CAG-Gs

WEAK CONSENSUS PATH EXISTS (WPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a weak consensus path from t?

A deterministic algorithm would iterate through each composite action $c \in E$ labelling those transitions which correspond to weak composite consensus actions. This is an $O(n \times |E|)$ operation (from WCA-C). After this pre-computation we apply algorithm 6.1 ($O(|T| \times |E|)$) to compute the set of states that are on consensus paths. Finally we verify that the given state t is found in CP ($O(|T|)$).

The overall time complexity of the deterministic algorithm is $O((n \times |E|) + (|T| \times |E|))$. The algorithm is polynomial in the number of agents, states, and transitions.

C-WEAK CONSENSUS PATH EXISTS (CM-WPE)

Given a T-CAG-G $\Gamma = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ and a state, $t \in T$, does there exist a C-minimal weak consensus path from t?

A deterministic algorithm would iterate through each composite action $c \in E$ labelling those transitions which correspond to C-minimal weak composite consensus actions. This is an $O((n \times |J|) \times |E|)$ operation (from CM-WCA-C). After this pre-computation we apply algorithm 6.1 ($O(|T| \times |E|)$) to compute the set of states that are on consensus paths. Finally we verify that the given state t is found in CP ($O(|T|)$).

The overall time complexity of the deterministic algorithm is $O(((n \times |J|) \times |E|) + (|T| \times |E|))$. The algorithm is polynomial in the number of agents, states, joint actions and transitions.

6.5 Discussion and Related Work

In this chapter we have presented temporal consensus action games with goals (T-CAG-G); these extend the CAG-G model by the inclusion of a temporal element, hence, T-CAG-Gs describe games that are played repeatedly, over time. Our approach to extending CAG-Gs to T-CAG-Gs differs significantly from that taken in previous chapters where, typically, we have introduced new structures to an extant model. Rather, we add a temporal element to CAG-Gs by embedding the CAG-G model within a transition system.

Specifically, a T-CAG-G is a Kripke structure in which each state $t \in T$ of the transition system is labelled with a CAG-G ($\gamma(t)$) and series of propositions representing both goals (G) and arbitrary atoms (P) with valuations $\pi_G(t)$ and $\pi_P(t)$, respectively. Hence, each state in a T-CAG-G fully describes not only the CAG-G to be played but also the goals presently achieved and an arbitrary world-state. Each transition in a T-CAG-G represents a composite action - one or more joint actions performed by disjoint coalitions of agents. By performing composite consensus actions the agents move from one state, t to another, t' and so from one CAG-G, $\gamma(t)$, to another, $\gamma(t')$.

We have defined temporal variants of solution concepts developed in earlier chapters. These consensus paths describe sequences of composite consensus actions. Since T-CAG-Gs are represented as Kripke structures the existence of our temporal solution concepts can be verified using established techniques from the field of model checking.

We have shown that the complexity of verifying the existence of consensus paths is polynomial in the number of agents, states and transitions for individually rational solution concepts, and in the number of agents, states, transitions and joint actions for collectively rational solution concepts. Whilst polynomial-time results are desirable we note that the size of these structures may be large and, in the worst case may themselves be exponential in the number of agents.

Multi-agent systems are rarely characterised as static, one-shot, affairs. More commonly a multi-agent system is viewed as a number of dynamic processes which persist over time. It is typical that such processes are situated within an environment that, itself, progresses and evolves. Given the centrality of coalitional problems within multi-agent research it is surprising that relatively little attention has been given to such problems in their repeated form.

There has been some interest in the specification of temporal logics for reasoning about coalitional behaviours in multi-agent systems. Such work includes coalition logic [Pauly, 2002] and alternating time temporal logic [Alur et al., 2002]. However, the principle focus of research in this area has been to formalise properties concerning, for example, the states of a system that a particular coalition may enforce. Typically these logics abstract away from the actual procedures of game-play, hence they have little to say about how, or which coalitions will form. A notable exception to this trend is coalitional game logic (CGL) [Ågotnes et al., 2006b]. CGL includes operators which permit representation of agents' preferences over outcomes. Ågotnes et al. show that CGL can provide logical characterisations for various game theoretic solution concepts including the core of non-transferable utility games.

T-CAG-Gs have some resemblance to the temporal qualitative coalitional games (TQCGs) of Ågotnes et al. [2006a]. TQCGs extend the QCG model of Wooldridge and Dunne [2004] by adding a temporal element to these games; in certain respects TQCGs are to QCGs as T-CAG-Gs are to CAG-Gs. As in T-CAG-Gs the TQCG model defines a set of states; similarly so, to each state is associated a game (QCG). However, our approach temporalisation differs from that of Ågotnes et al. in an important respect. We have chosen a branching temporal model for T-CAG-Gs, whereas Ågotnes et al. adopt a linear approach, associating to each time-point a state.

Whilst the linear model of time in TQCGs captures, perfectly, the notion of repeated game-play, i.e., that one game may be followed by another (potentially different) game, the ordering of games in TQCGs is deterministic: for a given time point the current state has exactly one successor. Hence, the temporal model found in TQCGs does not lend itself to modelling circumstances where games have multiple stable outcomes each suggestive of a different successor state. Our example (6.1) demonstrates one such circumstance, we argue that many naturally occurring problems fall into this category.

Chapter 7

Demonstration

7.1 Introduction

In this chapter we apply the T-CAG-G model introduced in Chapter 6 to a biologically inspired problem routinely faced by many social groups - that of maintaining spatial cohesion. In doing so we demonstrate the application of the model to a real-world problem and compare the predictions of our model with those found in the biological literature.

7.2 Background and Motivation

Group living is common place across a wide range of species. The success of this strategy has been attributed to such factors as decreased risk of predation [Hill and Lee, 1998] and increased efficacy through the pooling of information [Reebs, 2000]. For individuals to enjoy the benefits associated with a social existence it is necessary that the group remains spatially cohesive.

Biomechanical and physiological differences between individuals, brought about by e.g., variations in age, gender, and reproductive status can lead to divergent motivations and conflicting needs amongst group members [Leca et al., 2003]. Since activities relating to essential needs are typically carried out at disparate locations, conflicts of need invariably lead to conflicts regarding the group's location. For example, thirsty individuals prefer the group to be at a watering hole, those that are hungry prefer the group to site at feeding grounds, whilst for those that are tired a location suitable for resting is preferable.

For the group to remain cohesive, some individuals must engage in activities that are individually sub-optimal. The benefits of group living therefore frequently come at a cost — maintaining spatial cohesion requires that individuals synchronise the timing and location of their activities. Grouping costs increase as the preferences of individuals become more heterogeneous; the group becomes less synchronised with regard to the activity they would presently most prefer [Conradt and Roper, 2003].

Excessive costs cause the well being of individuals to suffer and can lead to the group fissioning, thus reducing grouping benefits for all group members. Decision making strategies employed by biological groups must therefore maximise the welfare of the group overall without excessively compromising the requirements of each individual [Conradt and Roper, 2007].

Theoretical and empirical research into group decision making strategies employed in biological groups suggests that both shared [Conradt and Roper, 2005] and unshared [King et al., 2008] decision making occur. In unshared decision making, decisions are made by a single individual within the group; such strategies are effectively dictatorships. In shared decision making each individual contributes to the decision making process.

7.3 Methodology

The use of theoretical models to study cohesion in natural groups under heterogeneity of preference is well established. Since the maintenance of spatial cohesion requires that certain group members compromise their optimal activity much of the prior work in this area has placed considerable emphasis on the modelling of individual costs and benefits. Such an approach inevitably leads to analysis through game-theoretic techniques [Conradt and Roper, 2007, 2009; DeDeo et al., 2010; Frank, 2009; Rands et al., 2008].

Our methodology is somewhat different. We define a problem in which a group of heterogeneous agents make repeated consensus decisions concerning their collective location; the problem is formally specified using the T-CAG-G model. We investigate the predictions of this model using model checking techniques typically employed in the verification of computational systems. Here, our interest is in the existence of and properties surrounding consensus paths.

To this end we created a bespoke software tool, implementing solutions for several decision problems considered earlier in this thesis. Specifically, we implemented algorithms for: strong consensus action (SCA, section 4.4.2), q_C-minimal strong consensus action (QM-SCA, also in section 4.4.2) and weak consensus action (WCA, section 4.4.3). For a given solution concept and T-CAG-G the tool labels those joint actions which are consensus actions. As described in sections 6.4.2 and 6.4.3 of this work such labelling is a prerequisite for computing the existence of a consensus path. The tool emits a representation of the labelled transition system in a form suitable for analysis using the PRISM model checker [Kwiatkowska et al., 2011]. Since PRISM is a probabilistic model checker we are not limited to merely establishing the existence of consensus paths; we also compute measures of likelihood concerning agents' behaviour with respect to these paths.

Model checking tools in general, and PRISM in particular have been used to investigate properties of a diverse range of computational systems. Amongst the previous uses of PRISM we find analysis of randomised distributed algorithms [Kwiatkowska et al., 2001, 2012; Ndukwu and McIver, 2010], network communication protocols [Duflot et al., 2006; Fehnker and Gao, 2006] and computer systems' security [Alexiou et al., 2010; Steel, 2006]. Interestingly, PRISM has also been used to study various biological processes including bone pathologies [Liò et al., 2011], cell signalling [Heath et al., 2008] and membrane fusion in viruses [Dobay et al., 2011].

A methodology such as ours has never before been used to investigate spatial cohesion in heterogeneous groups. Foremost, ours is the first treatment of this problem to consider the effects of endogenous thresholds upon group strategies for cohesion. Furthermore, ours is the first work to apply formal verification techniques to this field of research.

7.4 The Group Location Problem

In this section we define a biologically inspired problem in which a group of individuals make decisions regarding the location of the group. We assume that agents are situated in a simple environment which is divided into unique, discrete, interconnected locations. Agents may move about the environment without constraint and may remain at a location for as long as they wish. An example environment is illustrated in figure 7.1, transitions indicate the movement options available to agents in each location.

Each location is suitable only for a specific activity necessary for the agents' survival. Whilst away from a location an agent's need to be at that location increases; moving to the appropriate location satiates this need. Should an agent's need become too great then the agent expires.

At each location the group takes a decision regarding their next location. The group's movements form a route, or path through the environment. A solution to the group location problem is a path from location to location. We will say a solution is *successful* if consensus, and so cohesion, is maintained and all agents survive.

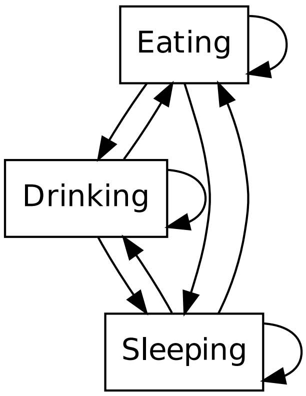

Figure 7.1: A simple environment in which agents may either eat, drink, or rest.

7.4.1 Formalisation of the group location problem

We now formalise the group location problem as a temporal consensus action game with goals (T-CAG-G). We refer to this formalisation as T-CAG-G(GLP).

We describe a problem instance for n agents $\{1, \ldots, n\}$, and $m, \{l_1, \ldots, l_m\}$ locations. Each agent has m goals, $\{g_1, \ldots, g_m\}$. Each location is suitable for the achievement of exactly one of the agents' goals and each location achieves a different goal.

In each location the grand coalition of agents may perform one of m joint actions, $j_1, \ldots, j_m$. Performing joint action j_k moves the grand coalition to location l_k where agents achieve the goal g_k.

For each goal, each agent $i \in A$ has a drive $\{d_{i,1}, \ldots, d_{i,m}\}$. Each drive assumes an integer value in the range $[0, \ldots, d_{\max}]$. The value of a drive represents the need of an agent to achieve the associated goal. The drive $d_{i,k}$ is (re)set to zero when agent i arrives at location k (and so achieves goal g_k). The value of drive $d_{i,k}$ is incremented at each move whilst the agent is not at location k. If, for any agent i $d_{i,k} = d_{\max}$ then that agent expires. Since drives are essentially counters and no two goals may be achieved simultaneously, the natural arrangement is that for any given agent no two drives may assume the same value. We will therefore limit our interest to those states where the drives of the agents are strictly ordered.

Formally, T-CAG-G(GLP) is the T-CAG-G
$\Gamma_{GLP} = \langle A, Ac, J, G, achieve, P, C, T, \gamma, \pi_G, \pi_P, \alpha, o\rangle$ where:

$A = \{1, \ldots, n\},\ n \geq 2$ a finite group of agents.

$Ac = \{a_1, \ldots, a_m\}$ is the set of individual actions.

$J = \{j_1, \ldots, j_m\}$ is a set of joint actions. Each $j_k \in J$ corresponds to every agent in the grand coalition performing the same action $a_k \in Ac$.

$G = \{g_1, \ldots, g_m\}$, the set of goals.

$achieve(j_k) = g_k$ for all $j_k \in J$.

P a set of propositions of the form $p_{i,k,x}$ for all $i \in A$, $1 \leq k \leq m$ and $0 \leq x \leq d_{max}$ representing the values of the agents' drives. For agent i the proposition $p_{i,k,x}$ is true only if for agent $i \in A$ drive $d_{i,k} = x$.

C a set of constraints of the form:

$p_{i,k,x} \iff \neg(\bigvee_{x' \neq x} p_{i,k,x'})$ — Agents' drives may take only a single value at any one time.

$p_{i,k,x} \iff \neg(\bigvee_{k' \neq k} p_{i,k',x})$ — Agents' drives are strictly ordered.

$g_k \iff \neg(\bigvee_{k' \neq k} g_{k'})$ — Only one goal may be achieved (true) at any one time.

T a set of states. The state $t \in T$ is determined by the propositions given in $G \cup P$ under the constraints of C.

$\gamma(t)$ labels each $t \in T$ with a CAG-G of the form $\langle A, Ac, J_t, G, goals_t, achieve, \{\succeq_{i,t} \mid i \in A\}, q_t\rangle$

$\pi_G(t) = \{g_k\}$ where g_k is the goal that is true in state $t \in T$

$\pi_P(t)$ is such that $p_{i,k,0} \in \pi_P(t)$ only if $\pi_G(t) = g_k$. Achieving goal g_k sets drive $d_{i,k} = 0$ for all agents $i \in A$.

$\alpha(t) = J$ for all $t \in T$ — All joint actions may be performed in every state.

$E_t = J$ for all $t \in T$ — All composite actions contain only a single joint action.

$o(t, c) \in T$ is such that for $c \in E_t$, where $c = j_k$, $\pi_G(o(t, c)) \supset achieve(j_k)$ and additionally for drives $d_{i,l}$, $l \neq k$, in state $o(t, c)$ these drives assume the value of $d_{i,l}$ in state t incremented by 1.

7.4.2 Specification of q for the group location problem

The selection of rational quorum values is constrained by the rationality postulates introduced in section 5.2 and the agents' orderings $\succeq_{i,t}$ over goal sets. For the group location problem we infer this ordering from the drive values of the agent. For the locations $k,\ l \in \{1, \ldots, m\}$ if $d_{i,k} \geq d_{i,l}$ we assume agent i weakly prefers location k to location l. Thus, for the corresponding joint actions $j_k,\ j_l \in J$, the goal achieved by j_k is weakly preferred to the goal achieved by j_l.

As quorum values are continuous the set of rational quorum functions is infinite. Many rational quorum functions will be equivalent in the sense that the agents will reach consensus (or otherwise) isomorphically; equivalent agents supporting at identical rounds. If j is a joint action for n agents then the canonical set of unique quorum functions is found where $q_t(i,\ j) \in \{0,\ \frac{1}{n},\ \ldots,\ \frac{n}{n}\}$.

Since the canonical set of unique quorum functions is discrete and bounded, for joint action $j \in J_t$ there are only a finite number of (combinations of) values for $q_t(i,\ j)$. A subset of these conform to rational behaviour. Without loss of generality we are going to consider only quorum functions for the group location problem that are rational, as specified by our rationality postulates, and that are canonical in form.

We identify states in which one or more agent has expired as those where $\exists i \in A$ such that for some $1 \leq k \leq m$, $d_{i,k} = d_{max}$. In these states further consensus action is not possible. Therefore, we require that for all states $t \in T$ where one or more agent has expired $q_t(i, j) = 1$ for all $i \in A$ and all $j \in J$.

7.4.3 Consensus paths and solutions in T-CAG-G(GLP)

We identify three distinct classes of solutions for the group location problem. The first corresponds to a successful solution where, to an infinite horizon, the agents maintain consensus, so remain cohesive, and no agent expires. Such a solution would be where, for example, the agents cyclically visit each location in turn, *ad infinitum*.

The second and third classes of solution correspond to non-successful solutions where the agents perform only consensus actions but they reach a state where further consensus action is not possible. Such *deadlock* states occur because either:

- one or more agent has expired, or,
- no agent has expired, but the agents cannot reach consensus regarding their next action (the group fissions).

Recall that a consensus path is a full path consisting only of consensus actions. Hence, for an instance of T-CAG-G(GLP) the agents can (potentially) be successful if and only if a consensus path exists from a given initial state. However, the existence of a consensus path does not guarantee that the agents will be successful. A state, belonging to a consensus path, may also present consensus actions which lead to states that do not belong to a consensus path. Thus, agents may deviate from the consensus path. Such deviations will ultimately lead to deadlock states.

7.5 Investigating T-CAG-G(GLP)

We are going to use formal model checking tools to investigate properties of the group location problem as described by the T-CAG-G(GLP) model. We are interested in identifying specifications of the quorum function that lead to the existence of consensus paths and hence successful solutions under the temporal solution concepts introduced in Chapter 6. We begin by observing that since, in T-CAG-G(GLP) all joint actions are for the grand coalition; the C-weak consensus path is not an informative solution concept for this model (since subsets of agents cannot act independently). Therefore we shall focus our attention on the existence of:

- strong consensus paths,
- weak consensus paths,
- q_C-strong consensus paths.

As we have noted, the existence of a consensus path is necessary for the agents to be successful, but is not sufficient to guarantee success for the agents. Hence, we are particularly interested in specifications for the quorum function that lead to certain success. Such cases are identified as those where, in every state belonging to a consensus path, all consensus actions (equivalently, all out-going transitions) lead to states that also belong to a consensus path. We shall refer to the set of all consensus actions in the model as the *consensus action space*. Where the consensus action space contains only consensus actions found on a consensus path we shall say that the consensus action space is *tight*.

We are also interested in comparing the behaviour of agents in the T-CAG-G(GLP) model with that of individuals in naturally occurring groups. Biological evidence suggests that the degree to which the agents are synchronised will have an effect upon how easily the agents can reach consensus. Establishing the degree to which natural individuals are synchronised is non-trivial since it requires that the internal state of each individual is known. However, since in T-CAG-G(GLP), we model the drives of the agents establishing a measure of synchronisation between agents is relatively straightforward. The constraints structure of T-CAG-G(GLP) is such that once the agents have visited each location once their drives will be synchronised. Thus, we are going to look at the effects of synchronisation in the *initial state* upon the existence of consensus paths and the tightness of the consensus action space.

Biological research also suggests that both shared and unshared decision making occur within natural groups. An interesting case of shared decision making under our solution concepts is found where all agents are homogeneous with respect to their implementation of the rationality postulates. We are interested in understanding whether agents that are homogeneous in this sense can be successful. In general the solution concepts we have considered are representative of a shared decision making strategy; each agent contributes to the eventual decision outcome. However, we are going to establish a set of specifications for quorum functions which correspond to the presence of a dictator within the group of agents. Using the model we will examine the performance of the agents where their collective actions are determined dictatorially.

Lastly, we take the opportunity to test an intuitive hypothesis regarding the mean quorum value of the agents for each joint action. Our motivation for investigating the mean quorum value begins with the intuition that a lower mean quorum value for a joint action should suggest that the joint action is more likely to be a consensus joint action.

Taking strong consensus as an example: where consensus is achieved in a single round we know that the mean quorum value will be zero (lower bound). Similarly, the mean has an upper bound beyond which consensus cannot be reached; the mean reaches its upper bound, $\left(\frac{n-1}{2}\right)^{-1}$, where consensus is reached in n rounds. The curiosity is that simple counter-examples show that this intuition is only partial, e.g., the mean of $\{0.1, 0.1, 0.1, 0.1\}$ is less than the mean of $\{0, 0.25, 0.5, 0.75\}$ yet the second is a strong consensus action whilst the first is not. We are interested in identifying, for each solution concept, whether there exists a relationship between the mean quorum value and the likelihood of success.

In the remainder of this section we formalise our investigation of the T-CAG-G(GLP) and describe our experimental design.

7.5.1 Model parameters

The T-CAG-G(GLP) is parameterised by the following variables:

- n — the number of agents,
- m — the number of locations,
- d_{max} — the maximum value for the agents' drives.

Since we are using model checking tools to verify properties of the model and such techniques can be computationally expensive we consider instances of a relatively small size. Therefore we shall consider instances of T-CAG-G(GLP) for $n = 3$ agents and $m = 3$ locations.

We are now going to consider the value for the model parameter d_{max}. This parameter determines how needful agents may become before they expire. Since d_{max} places an upper bound on the value of each drive, the size of the state space for T-CAG-G(GLP) is exponential in d_{max}. For obvious reasons we want to avoid state-bloat so we would like to keep d_{max} as small as possible. To minimise d_{max} we will assume that each drive assumes the minimum value necessary to meet the model constraints. Thus, in the initial state, each agent will have one drive, the drive that is most in need of satiation, with a value of two. We will refer to this drive as the *most urgent* drive.

If d_{max} is too small then it may not be possible for the agents to be successful at all. For example, if $d_{max} = 4$ then, from initial states where agents differ in their most urgent drive it is not possible for the agents to visit each location even once. The most urgent drive of some agent will be sated on the first move, but in the subsequent state the most urgent drives of the remaining two agents will now be at d_{max}. Whichever of the remaining agents is sated in the following state, the other agent will expire. This does not seem 'fair'.

We choose the minimum fair value for d_{max} to be $d_{max} = 5$. This value is fair in the sense that even where the agents' most urgent drives are initially in opposition it is possible for every agent to satiate every goal once before it expires.

7.5.2 Strategies and specification of quorum functions

In this section we describe our approach to the specification of rational quorum functions for T-CAG-G(GLP). Recall that for this model the ordering relation of agent $i \in A$, $\succeq_i$, is determined as a function of the drives of agent i given the state $t \in T$. Recall also that the values of these drives, for each agent and in all states, are strictly ordered. Thus where $m = 3$, in each state, each agent will have exactly one drive that is most urgent (corresponding to the agent's non-dominated goal), one drive that is less urgent and one drive (corresponding to the most recently achieved goal) that is least urgent.

In state $t \in T$, for agent $i \in A$ and joint action $j_k \in J$ where $achieve_i(j_k) \in max_i$ (i.e. g_k is non-dominated under $\succeq_i$) the postulate of unconditional support requires that $q_t(i, j) = 0$. Therefore agents unconditionally support the joint action that achieves the goal corresponding to their most urgent drive. Quorum values for the remaining joint actions, corresponding to the lesser and least urgent drives must then be selected according to the postulate of proportional support.

Formally, we shall refer to the joint action achieving the agent's non-dominated goal as j_d, the joint action achieving the goal ranked second-most as j_s and the joint action achieving the goal ranked third-most (least) as j_t. Then, for each agent $i \in A$ let $\sigma_i = \langle q_t(i, j_d), q_t(i, j_s), q_t(i, j_t)\rangle$ be a tuple of quorum values. We refer to σ_i as the *strategy* of i and write $\sigma_i = k$ to indicate that agent $i \in A$ is following the strategy numbered k (see table 7.1 for the range of values for k) . We refer to the collective strategies of all agents as the *group strategy* and write $\sigma_A = \langle \sigma_1, \sigma_2, \sigma_3\rangle$ to indicate the individual strategies of each agent .

For each model instance, each agent $i \in A$ is assigned a strategy, σ_i. The strategy defines the set of quorum values the agent will use for this instance. The group strategy profile σ_A fully prescribes the quorum function q_t for each state in the game. In each strategy $q\#(i, j_d) = 0$ for the joint action, j_d which achieves the agent's non-dominated goal-set. For the remaining $m - 1 = 2$ joint actions we have j_s achieving the goal ranked second and j_t achieving the goal ranked third, $q\#(i, j_s), q\#(i, j_t) \in \{0, 1, 2, 3\}$.

Where $n = m = 3$ each agent may follow one of exactly 10 rational strategies. Table 7.1 shows the rational strategy profiles for a single agent $i \in A$; for simplicity the table shows $q\#$ values; i.e., exact numbers of agents.

Table 7.1: Rational strategy profiles for $n = m = 3$

σ_i	$q\#(i, j_d)$	$q\#(i, j_s)$	$q\#(i, j_t)$
0	0	0	0
1	0	0	1
2	0	0	2
3	0	0	3
4	0	1	1
5	0	1	2
6	0	1	3
7	0	2	2
8	0	2	3
9	0	3	3

The problem is sufficiently small that we can consider all 10^3 combinations of group strategies. We will therefore exhaustively enumerate the entire space of rational, canonical quorum functions.

There are two classes of group strategy which are of interest. These are homogeneous group strategies in which all agents follow the same strategy, and dictatorship strategies. We say a group strategy is homogeneous where $\forall i \in A, \sigma_i = s$, where $0 \leq s \leq 9$.

We say a group strategy is a dictatorship where exactly one agent $i \in A$ has the strategy $\sigma_i = 9$. The behaviour of agents following this strategy is to unconditionally support the joint action that achieves their non-dominated goal and to object to all other joint actions. Such agents function as dictators since they will not engage in activities which, for them, are sub-optimal.

7.5.3 Synchronisation and initial states

In T-CAG-G(GLP) the state representation describes the location of the group (the goal that is currently achieved) and the drives of each agent. Since the environment is fully connected, in the sense that every location is reachable from every other location the choice of initial location is arbitrary. Our interest is in the effects of heterogeneity of preference upon outcomes for the group; therefore we are going to focus on the ordering of the drives of the agents in the initial state. This ordering determines the ordering of agents with respect to their goals and so determines the degree of conflict initially present in the group.

We want to establish a small set of suitable initial states. These states will be used as the basis for considering the effects of heterogeneous preferences upon the group's ability to repeatedly reach consensus and be successful. Where two agents have the same ordering over locations we will say that the agents are *synchronised*; for agents $i, i' \in A$, and with slight abuse of notation, we write $\succ_i = \succ_i'$. Increasing the number of synchronised agents is one way to systematically control the heterogeneity of preferences within the group. We can now characterise three kinds of initial state.

- We will say the agents are *fully synchronised* in the initial state if $\forall i, i' \in A \succ_i = \succ_{i'}$; all agents have the same ordering over drives.
- We will say the agents are *partially synchronised* in the initial state if $\exists i, i' \in A, i \neq i'$ such that $\succ_i = \succ_{i'}$ and $\exists i'' \in A \setminus \{i, i'\}$ such that $\succ_{i''} \neq \succ_i$; two agents have the same ordering over drives whilst the remaining agent's ordering differs.
- We will say the agents are *not synchronised* in the initial state if $\forall i, i' \in A, i \neq i', \succ_i \neq \succ_{i'}$, no two agents have the same ordering over drives.

Given the symmetries about agents, drives and locations which are present in the group location problem we need only consider a single ordering for the case where the agents are fully synchronised. Cases where synchronisation is partial or non-existent possess several possible orderings. To consider these we shall fix the drives of a single agent to constant values whilst exploring all possible combinations of drive values for the remaining agents. Given the symmetries and constraints that exist within the problem this approach is exhaustive.

In total we shall consider 36 configurations of the agents' drives in the initial state. In addition to the single initial state where the agents' drives are fully synchronised we consider 15 configurations where the drives are partly synchronised and 20 where the drives are not synchronised. Complete details of the drives values for agents in each initial state can be found in table 1 of Appendix A.

7.5.4 Experimental variables

We can now state our experimental variables. For our investigation of T-CAG-G(GLP) the independent variables describe the modality of consensus, the specification of the quorum function and the degree to which the agents are synchronised: These are:

- the temporal solution concept,
- the group strategy of the agents,
- the initial state and its kind.

Our dependent variables measure the existence of a consensus path and how challenging it is for the agents to repeatedly perform consensus actions, and so follow a consensus path and be successful. Given a solution concept, a group strategy and initial state, the dependant variables are:

- whether a consensus path exists,
- the minimum probability of consensus action in a successor state,
- the average probability of consensus action in a successor state.

We give precise details of how the dependent variables are computed in the following section.

7.6 Experimental Method

The group strategy of the agents and their initial state combine to uniquely describe one *instance* of T-CAG-G(GLP). We exhaustively enumerate all 1,000 group strategies and 36 initial states giving a total of 36,000 instances of T-CAG-G(GLP). For each solution concept (strong, weak and q_C-minimal) we used a bespoke software tool to encode the consensus action space as a discrete time Markov chain (DTMC) model.

Each state in the DTMC model corresponds to one state, $t \in T$, in T-CAG-G(GLP). Each transition in the DTMC model corresponds to a consensus action under the given solution concept. A state in the DTMC is a deadlock state if an agent has expired, or the agents cannot reach consensus regarding any joint action. Since we are interested in three solution concepts (strong consensus, weak consensus and q_C-minimal strong consensus) in total we encoded 108,000 DTMC models.

Each model was encoded using the Reactive Modules state-based language of the probabilistic model checking tool PRISM [Kwiatkowska et al., 2011]. To compute the independent variables we then verified the following properties:

1. *Does a consensus path exist from the initial state?*

   Property: $E[G!``deadlock"]$

   Meaning: Does there exist a full path from the initial state where globally, no states are deadlock states. This property is true only if a consensus path exists and false otherwise.

2. *How challenging, in the worst case, is it for the agents to keep performing consensus actions?*

   Property: $filter(min, P =?[X!``deadlock"], !``deadlock")$

   Meaning: Considering all non-deadlock states reachable from the initial state what is the minimum probability that the next state on a path from a non-deadlock state is also a non-deadlock state. The property reports, over all reachable non-deadlock states, the minimum probability that a non-deterministically selected consensus action will lead to a state with at least one outgoing transition corresponding to a consensus action. We denote this property as $p(minXca)$

3. *How challenging on average is it for the agents to keep performing consensus actions?*

   Property: $filter(avg, P =?[X!``deadlock"], !``deadlock")$

   Meaning: Considering all non-deadlock states reachable from the initial state what is the average probability that the next state on a path from a non-deadlock state is also a non-deadlock state. The property reports the average (mean over all reachable non-deadlock states) probability that a non-deterministically selected consensus action will lead to a state with at least one outgoing transition corresponding to a consensus action. We denote this property as $p(avgXca)$.

Where $p(minXca) = 1$ the consensus action space contains only consensus actions which lie on a consensus path, the consensus action space is tight. Where $p(minXca) < 1$ then the consensus action space contains at least one consensus action that does not exist on a consensus path. A value of e.g., 0.333 indicates that there exists at least one reachable non-deadlock state where the probability of non-deterministically selecting a consensus action leading to a state where further consensus action is possible is as low as one third. The property $p(avgXca)$ can be interpreted in an similar manner, $p(avgXca) = 1$ if and only if $p(minXca) = 1$; the consensus action space is tight.

Whilst the $p(minXca)$ property considers the worst case (found in as few as a single state) the $p(avgXca)$ property is representative of the consensus action space as a whole. Together these properties characterise how challenging it is for the agents to repeatedly perform consensus actions, and so follow a consensus path and be successful. As these probabilities increase the agents are more likely to follow a consensus path if they (non-deterministically) perform only consensus actions. Therefore these measures are indicative of the *tightness* of the consensus action space.

7.7 Results

In this section we give our experimental results. We begin with general results concerning the existence of consensus paths and tightness of the consensus action space; we then turn to the effects of the initial state upon these. Following this we present results regarding the performance of group strategies and in particular those strategies that are homogeneous and those where the group contains a dictator. Finally we examine the relationship between the mean quorum value and the existence of consensus paths and tightness of the consensus action space. In each case we present results for each of the three relevant solution concepts.

7.7.1 Consensus paths

Consensus Path Existence

We looked at consensus path existence for each solution concept over all initial states and strategies. For the 36,000 instances of T-CAG-G(GLP) table 7.2 shows, for each solution concept, the proportion of these instances where a consensus path exists.

Table 7.2: Proportion of instances where a consensus path exists.

Solution concept	Proportion
Strong consensus path	0.8915
Weak consensus path	0.9190
q_C-Strong consensus path	0.8915

We found that consensus paths exist in a high proportion (approximately 90%) of instances for all solution concepts. The results for strong consensus paths, and their collectively rational counterpart q_C-strong consensus paths are identical since, by definition, wherever a strong consensus path exists a q_C-strong consensus path also exists. The proportion of instances where a consensus path exists is highest for weak consensus. This reflects that all strong consensus actions are also weak consensus actions therefore all strong consensus paths are also weak consensus paths. We found that for a small (2.75%) proportion of instances weak consensus paths exist where strong consensus paths do not. Since the weak consensus action space is a super set of the strong consensus action space, which itself is a superset of the q_C-minimal consensus action space the results tell us that in just over 8% of instances no consensus path (of any kind) exists.

Following Consensus Paths

Here we give results concerning how challenging it is for the agents to follow consensus paths under our solution concepts. For each solution concept we computed the means of properties $p(minXca)$ and $p(avgXca)$, $\overline{p(minXca)}$ and $\overline{p(avgXca)}$, respectively. These are measures of how challenging it is for the agents to follow a consensus path. Table 7.3 shows the mean values of these properties. The first pair of columns show means over all 36,0000 instances, the second pair show means only where a consensus path exists.

Table 7.3: Means of $p(minXca)$ and $p(avgXca)$ for each solution concept.

	All instances		**Consensus path exists**	
Solution concept	$\overline{p(minXca)}$	$\overline{p(avgXca)}$	$\overline{p(minXca)}$	$\overline{p(avgXca)}$
Strong consensus path	0.5852	0.7773	0.6564	0.8720
Weak consensus path	0.4515	0.7449	0.4913	0.8105
q_C-Strong consensus path	0.8567	0.8796	0.9610	0.9866

Although weak consensus paths exist in the greatest proportion of instances we find that weak consensus paths are the hardest to follow. These results are consistent and tell us that whilst weak consensus is reached in a wider array of circumstances than strong consensus there are many consensus actions in the weak consensus action space which do not lie on a weak consensus path. Interestingly, q_C-strong consensus paths are the easiest for the agents to follow. Where a q_C-strong consensus path exists, the average probability that the agents will choose a consensus action leading to a non-deadlocked state is almost 99%.

We now look at instances where the consensus action space is tight; these are identified by $p(minXca) = p(avgXca) = 1$. Table 7.4 shows the proportion of instances where the consensus action space is tight; proportions are given as before, for all instances and for only those where a consensus path exists.

Table 7.4: Proportion of instances where the consensus action space is tight.

Solution concept	**All instances**	**Consensus path exists**
Strong consensus path	0.3417	0.3833
Weak consensus path	0.1960	0.2133
q_C-Strong consensus path	0.8252	0.9257

The results show that in T-CAG-G(GLP) the weak consensus solution concept yields a tight consensus action space on far fewer occasions than strong or q_C-minimal consensus. Moving from weak consensus to strong consensus and from strong consensus to q_C-strong consensus the consensus action space not only becomes smaller but also is more likely to be tight. In particular we found the q_C-minimal consensus action space was tight in over 80% of all instances; this rises to over 90% if we consider only instances where a q_C-strong consensus path exists. For T-CAG-G(GLP) this means that the agents are most likely to follow a consensus path and so be successful if the q_C-strong consensus solution concept is applied.

Where a weak consensus path exists the consensus action space is tight in only a little over 20% of instances. Since weak consensus can be reached in situations where strong consensus cannot, this raises the question of whether the weak consensus action space can be tight where strong consensus paths are not found. We looked at those instances where a weak consensus path exists but a strong consensus (equivalently q_C-strong consensus path) path does not exist. Of these we found that in just over one third (0.3508) of instances where a weak consensus path exists the weak consensus action space is tight.

7.7.2 Initial states

Here we give results concerning the effects of the initial state kind upon the existence of consensus paths and tightness of the consensus action space. The 36,000 instances are broken down by initial state kind as follows:

- 1,000 where the agents are fully synchronised in the initial state (Full)
- 15,000 where the agents are partly synchronised in the initial state (Part)
- 20,000 where the agents are not synchronised in the initial state (None)

Effects of initial state kind upon existence of consensus paths

For each initial state kind we computed the proportion of that state kind for which a consensus path exists. Table 7.5 gives these results broken down by solution concept.

Table 7.5: Existence of consensus paths given the synchronisation in the initial state.

Solution Concept	**Synchronisation**	**Proportion**
Strong consensus path	Full	1.0000
Weak consensus path	Full	1.0000
q_C-Strong consensus path	Full	1.0000
Strong consensus path	Part	0.9206
Weak consensus path	Part	0.9402
q_C-Strong consensus path	Part	0.9206
Strong consensus path	None	0.8642
Weak consensus path	None	0.8991
q_C-Strong consensus path	None	0.8642

These results show that the initial state kind has a clear effect upon the existence of consensus paths across all solution concepts. Where the agents are fully synchronised a consensus path always exists, this is to be expected since there is no conflict of interest between the agents. As the agents become less synchronised in the initial state the proportion of instances where a consensus path exists decreases.

We found that the proportion of instances where a consensus path exists was at its lowest, for all solution concepts, where the agents are not synchronised in the initial state. This proportion reached a minimum, of only 80%, in initial states numbered #8 and #13 (see table 1 of Appendix A for further details of these states). In particular these two initial states can be distinguished as the only initial states in which the agents' ordering over their goals are rotations of each other.

Effects of initial state kind upon following consensus paths

We now give results concerning the effects of the initial state kind upon the tightness of the consensus action space. For each solution concept and initial state kind, table 7.6 shows the means for $p(minXca)$ and $p(avgXca)$ over all instances and for only those where a consensus path exists. Note that the results for fully synchronised initial states states reflect that a consensus path always exists from such states.

Table 7.6: Means of $p(minXca)$ and $p(avgXca)$ for each solution concept and initial state kind.

		All instances		**Consensus path exists**	
Solution concept	**Sync.**	$\overline{p(minXca)}$	$\overline{p(avgXca)}$	$\overline{p(minXca)}$	$\overline{p(avgXca)}$
Strong consensus path	Full	0.7113	0.8960	0.7113	0.8960
Weak consensus path	Full	0.5995	0.8402	0.5995	0.8402
q_C-Strong consensus path	Full	0.9678	0.9892	0.9678	0.9892
Strong consensus path	Part	0.6106	0.8078	0.6633	0.8774
Weak consensus path	Part	0.4677	0.7680	0.4974	0.8169
q_C-Strong consensus path	Part	0.8861	0.9089	0.9625	0.9873
Strong consensus path	None	0.5598	0.7486	0.6478	0.8662
Weak consensus path	None	0.4320	0.7227	0.4805	0.8038
q_C-Strong consensus path	None	0.8291	0.8520	0.9594	0.9859

Table 7.7 shows the proportion of instances from each initial state kind where the consensus action space was tight. As before, proportions are given out of the total

number of instances with each initial state kind and the also for those only where a consensus path exists.

Table 7.7: For each initial state kind, the proportion of instances where the consensus action space is tight.

Solution concept	Sync.	All instances	Consensus path exists
Strong consensus path	Full	0.4440	0.4440
Weak consensus path	Full	0.2710	0.2710
q_C-Strong consensus path	Full	0.9360	0.9360
Strong consensus path	Part	0.3646	0.3960
Weak consensus path	Part	0.2112	0.2246
q_C-Strong consensus path	Part	0.8546	0.9283
Strong consensus path	None	0.3194	0.3696
Weak consensus path	None	0.1809	0.2012
q_C-Strong consensus path	None	0.7976	0.9229

The results show that the initial state has a clear effect upon how easy it is for the agents to follow consensus paths. We found that, for all initial states kinds, the q_C-minimal consensus action space is most likely to be tight, hence q_C-strong consensus paths are the easiest to follow; then strong consensus paths and lastly weak consensus paths. However, all solution concepts exhibited a similar response to the degree of synchronisation amongst the agents. For all solution concepts it is easiest to follow consensus paths from a fully synchronised state, and hardest from a non synchronised state. Similarly, it is more likely that the consensus action space is tight where the agents are fully synchronised in the initial state, and least likely where the agents are not synchronised.

7.7.3 Strategies

Here we give results concerning the effects of group strategies upon outcomes for the agents. We exhaustively explored the space of rational quorum functions for T-CAG-G(GLP), in total we considered 1,000 group strategies. We begin by considering those group strategies for which a consensus path exists and the consensus action space is tight irrespective of initial state. We then present results concerning the effects of group strategies that are homogeneous and those where a dictator is present.

Resilience to the effects of initial state

For all solution concepts we found that a large number of group strategies were resilient to the effects of the initial state. For these group strategies a consensus path exists and the consensus action space can be tight across all of the initial states of the agents.

Table 7.8 shows the number of group strategies where consensus paths exist and also where the consensus action space is tight, irrespective of initial state. These results are given for each solution concept.

Table 7.8: For each solution concept the number of strategies which are resilient to the initial state.

Solution concept	**Consensus path exists**	**Tight consensus action space**
Strong consensus path	668	158
Weak consensus path	810	108
q_C-Strong consensus path	668	602

Of the 1,000 group strategies a large proportion, between approximately 70% to 80% (dependant on solution concept), showed consensus path existence for all 36 initial states. In particular we found for that for the q_C-strong consensus path solution concept just over 60% of group strategies gave rise to a tight consensus action space; this is considerably higher than for either weak or strong consensus.

Homogeneous Strategies

We looked at the effects of homogeneous strategies upon the existence of consensus paths and tightness of the consensus action space.

For each of the solution concepts and each homogeneous strategy we computed:

- the proportion of initial states where a consensus path exists (blue)
- the proportion of initial states where the consensus action space is tight (red)

Figure 7.2 shows these results for strong consensus and q_C-strong consensus (data are identical) whilst results for weak consensus are shown in Figure 7.3.

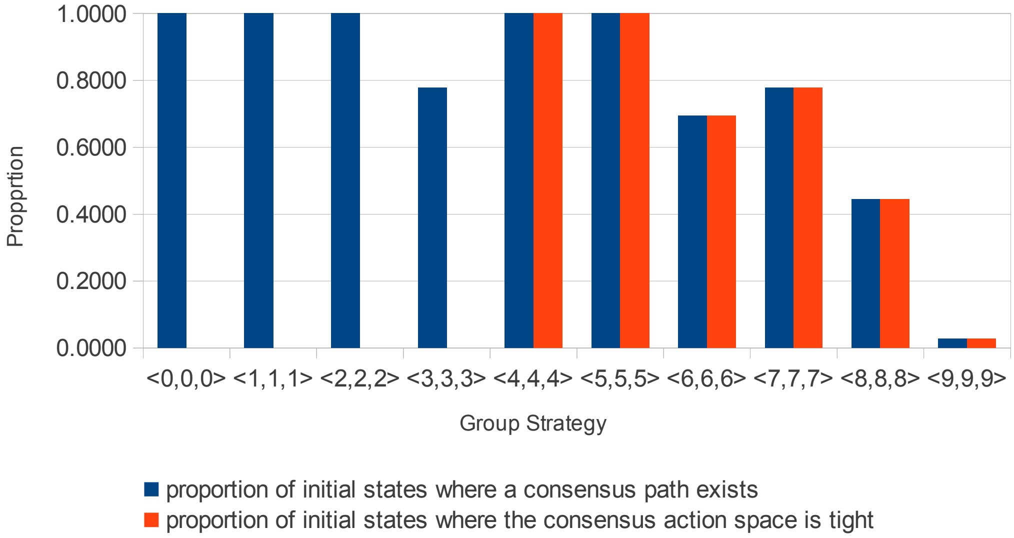

Figure 7.2: Effects of homogeneous strategies for strong and q_C-strong consensus paths.

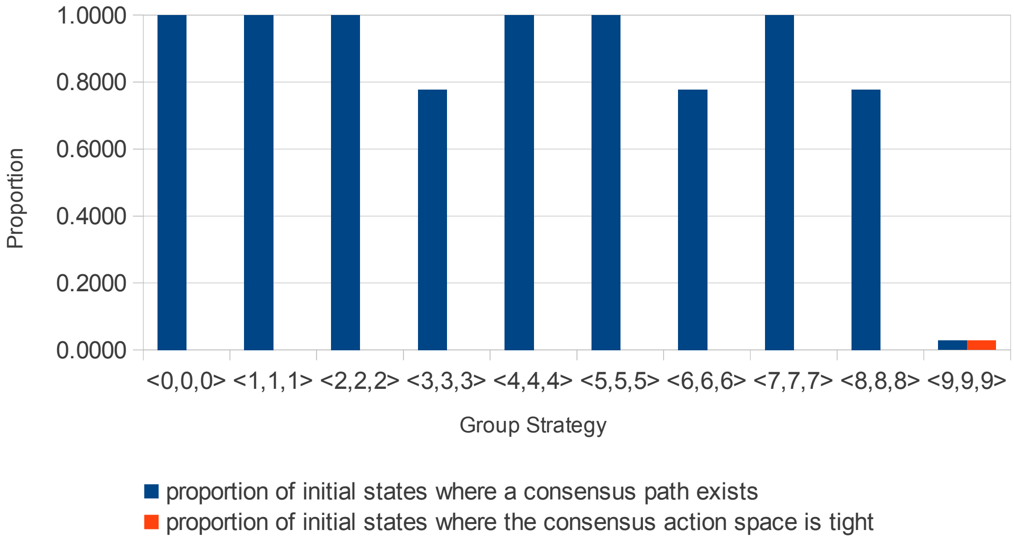

Figure 7.3: Effects of homogeneous strategies for weak consensus paths.

We found that consensus paths exist for all solution concepts for all homogeneous strategies. Figure 7.2 shows that for strong and q_C-strong consensus paths, where the group strategy, σ_A, is low numbered, between $\langle 0, 0, 0 \rangle$ and $\langle 3, 3, 3 \rangle$, although consensus paths exist the consensus action space is not tight. This reflects that in these strategies (see table 7.1) the agents unconditionally support the joint action that achieves their non-dominated goal (j_d) and also the joint action achieving the goal ranked second-most (j_s). Such agents can be described as too willing to agree, or too 'selfless'. Hence, for these strategies, although consensus paths exist for a large proportion (in many cases, all) of initial states, the consensus action space is never tight.

However, for all other homogeneous strategies figure 7.2 shows that whenever a consensus path exists the consensus action space is tight. Group strategies where $\sigma_A = \langle 4, 4, 4 \rangle$ or $\sigma_A = \langle 5, 5, 5 \rangle$ are particularly interesting. Consensus paths exist from all initial states for both of these homogeneous strategies; moreover, the consensus action space is tight irrespective of initial state. In these group strategies agents require at least one other agent to support the joint action achieving either their second or third-most ranked goals, but will never object to a joint action.

The strong/q_C-minimal consensus action space is also guaranteed to be tight for group strategies where σ_A lies between $\langle 6, 6, 6 \rangle$ and $\langle 9, 9, 9 \rangle$. However, in these strategies the agents become increasingly less inclined to perform anything other than the joint action j_d, achieving their non-dominated goal. For these group strategies the existence of a consensus path is not assured.

Figure 7.3 shows that for weak consensus the agents reach agreement in a wider range of circumstances than for strong consensus. Consequently the agents reach consensus over a larger set of joint actions. Therefore only in the 'most selfish' group strategy, $\sigma_A = \langle 9, 9, 9 \rangle$, where all agents object to joint actions other than j_d, is the consensus action space tight.

Dictatorships

Of the 1,000 group strategies there are 243 strategies that qualify as dictatorships having exactly one agent $i \in A$ where $\sigma_i = 9$. We looked at the effects of each of these dictatorial strategies in each of the 36 initial states (8,748 instances in total). For each of the solution concepts and each dictatorial strategy we computed:

- the proportion of initial states where a consensus path exists,
- the proportion of initial states where the consensus action space is tight.

For strong consensus paths and q_C-minimal consensus paths the proportion of initial states where a consensus path exists was 0.7503, of these, the consensus action space was tight in every case. For weak consensus paths the proportion of initial states where a consensus path exists was 0.7901, again the consensus action space was tight in every case.

We found that presence of a dictator is not sufficient to ensure the existence of a consensus path. However, where a consensus path does exist all dictatorial strategies give rise to a consensus action space that is tight; this is irrespective of solution concept. Consensus paths exist in slightly more (about 4%) cases for weak consensus than for strong consensus.

7.7.4 Mean quorum value

We asked if there is any correlation between the mean quorum value of a group strategy and:

- whether a consensus path exists,
- whether the consensus action space is tight.

For simplicity we are going to report $q\#$ values throughout this section — recall that these represent actual numbers of agents. We begin by looking at the frequency distribution of the mean quorum value over all instances. Figure 7.4 shows that for T-CAG-G(GLP) the mean quorum value assumes one of 19 distinct values and is normally distributed:

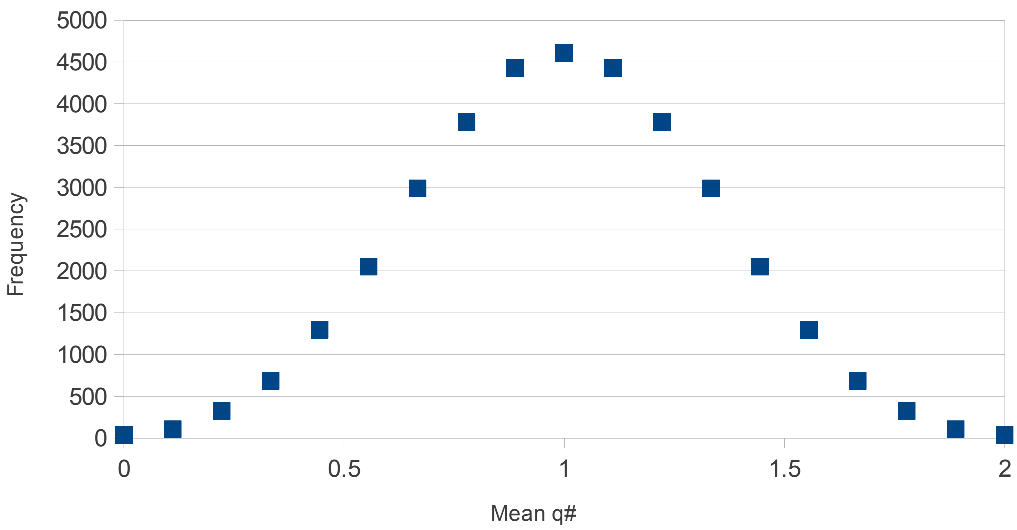

Figure 7.4: Frequency distribution of mean quorum values.

For each solution concept and each distinct value of the mean quorum value we computed the proportion of instances where a consensus path exists and where the consensus action space was tight. For example (for all solution concepts) there were 36 instances (one for each initial state) where the mean $q\#$ value was 0, of these all 36 (proportion is 1) exhibited a consensus path, but on no occasion (proportion is 0) was the consensus action space tight. These data are shown in figures 7.5 to 7.10.

Strong consensus path

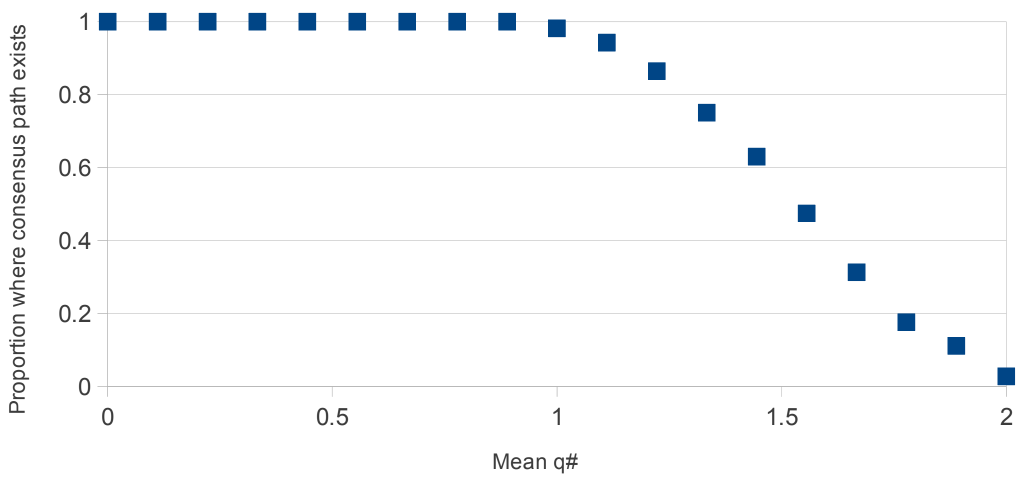

Figure 7.5: Mean quorum value vs. proportion of strong consensus path existence.

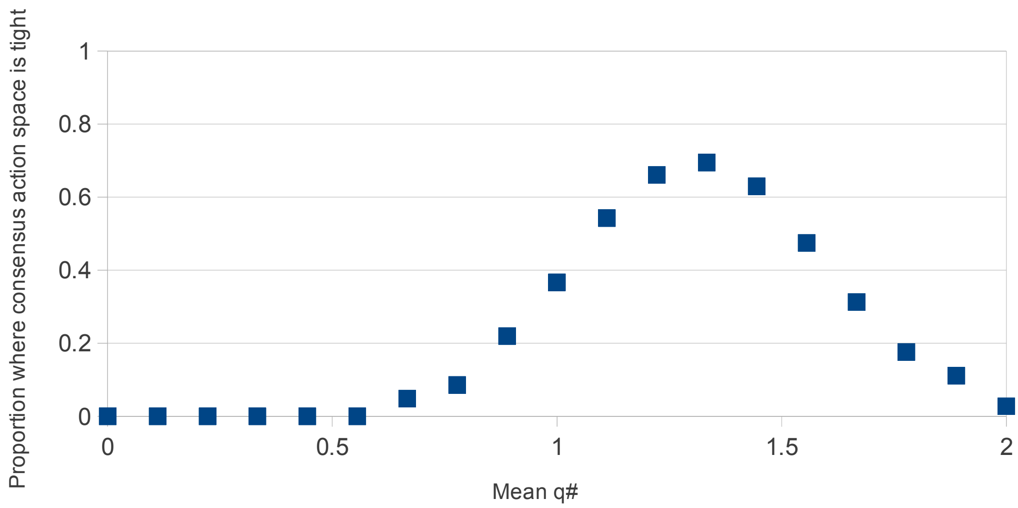

Figure 7.6: Mean quorum value vs. proportion of tight strong consensus action space.

Weak consensus path

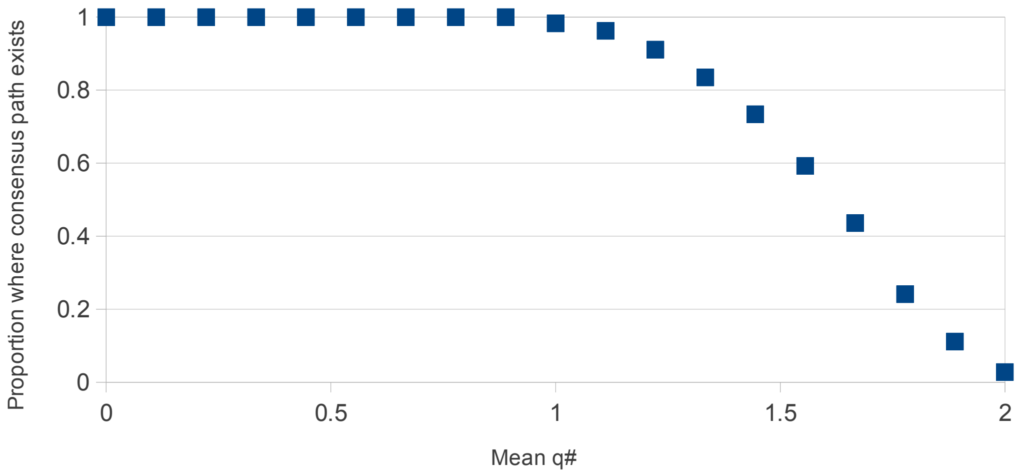

Figure 7.7: Mean quorum value vs. proportion of weak consensus path existence.

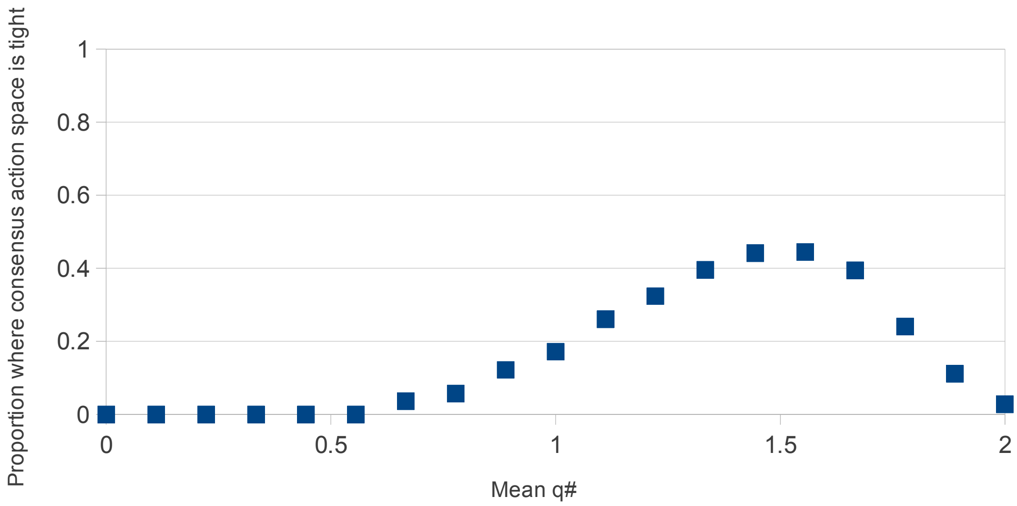

Figure 7.8: Mean quorum value vs. proportion of tight weak consensus action space.

q_c-Strong consensus path

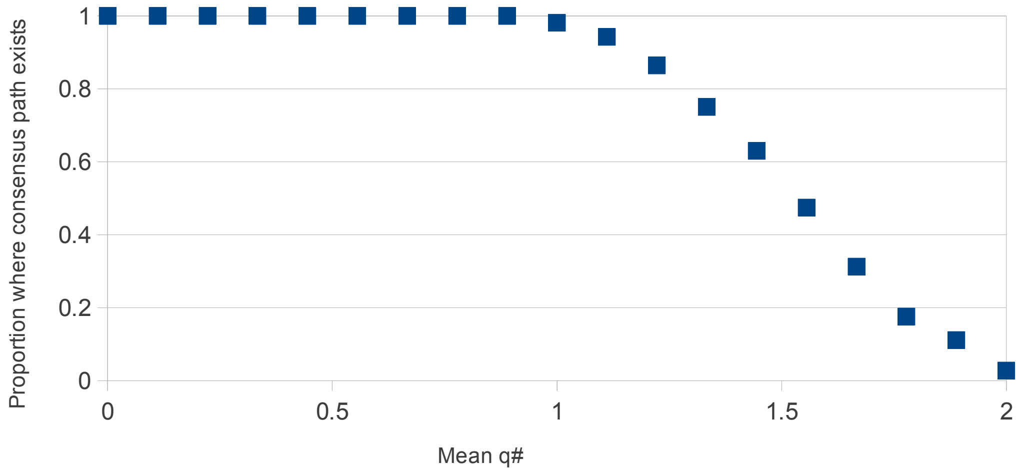

Figure 7.9: Mean quorum value vs. proportion of q_C-strong consensus path existence.

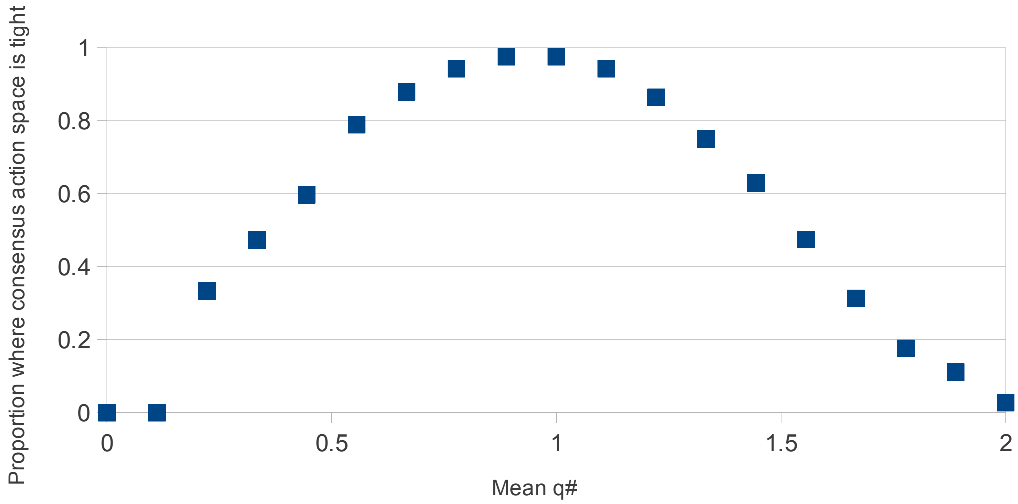

Figure 7.10: Mean quorum value vs. proportion of tight q_C-minimal strong consensus action space.

The plots suggest that the mean quorum value and the existence of consensus paths/tightness of consensus action space are related for all three solution concepts. However these relationships are clearly non-linear. For all solution concepts increasing the mean $q\#$ value beyond one agent decreases the probability that a consensus path will exist. This is consistent with the intuition that a higher mean corresponds to a lower probability of consensus. An interesting question, therefore, is whether the mean $q\#$ is a significant predictor of the existence of consensus paths/tightness of consensus action space.

Is the mean $q\#$ value a significant predictor?

To understand whether the mean quorum value is a significant predictor of the existence of a consensus path/tight consensus action space we performed a binary logistic regression (generalised linear model with logit link function). A post-hoc power analysis was undertaken to confirm that our sample size of 36,000 is sufficient for analysis at the 0.99 confidence interval.

We performed six logistic regression analyses. In each we used the mean $q\#$ value as the independent variable. The six covariates were dichotomous variables representing the existence of a consensus path and whether consensus action space was tight, for each of the three solution concepts. The existence of a tight consensus action space was coded as a dummy variable assuming the value 1 (true) iff $p(minXca) = p(avgXca) = 1$.

In each case we found the mean $q\#$ value to be a significant ($p < 0.0001$) predictor of the covariate. To understand the predictive power of the mean q# value we looked at the Nagelkerke R Square value. This can be interpreted as the proportion of the variance in the covariate that can be explained by the independent. These data are shown in the table 7.9.

Table 7.9: Nagelkerke R Square for mean quorum value as a predictor of consensus path existence and tight consensus action space.

Solution concept	**Consensus path exists**	**Tight consensus action space**
Strong consensus path	0.470	0.245
Weak consensus path	0.432	0.186
q_C-Strong consensus path	0.470	0.240

These results show that although there is a significant relationship between the mean quorum value and the existence of consensus paths, the mean quorum value is not a robust predictor of the existence of a consensus path, accounting for less than 50% of the variance in the output of the model. This result reflects the high variance in our data due to factors such as the choice of initial state. Similarly our results also suggest that the mean quorum value is also not a robust predictor of whether the consensus action space is tight. Overall these results are consistent with our intuition that although a lower mean quorum value increases the probability of consensus, in general consensus can be reached across a broad range of the mean quorum value.

7.8 Discussion and Related Work

We have introduced and studied the group location problem in which a group of agents repeatedly achieve individual goals whilst remaining spatially cohesive. Cohesion is achieved by maintaining consensus over repeated decisions regarding the location of the group. The problem is challenging since the agents may be in conflict regarding their goals and hence the next location of the group. Our work is the first to consider collective selection of group location where agents' choices are mediated by individually rational thresholds. Ours is also the first work to examine such biological problems using formal verification and model checking techniques.

However, ours is not the first application of agent-based techniques to the study of movements in natural groups. Sellers et al. [2007] uses agent-based simulation to model group movement in a complex, dynamic and realistic environment where agents' energetic requirements are governed by bio-realistic functions and group decisions regarding whether or not the group should move are made by voting with a variable, majority threshold. They found that the majority threshold is a strong predictor of whether or not the agents will achieve their goals. However their work does not consider agents with individual thresholds, or the effects of the initial state of the agents. Moreover, in their model spatial cohesion, hence consensus, is enforced, even where agents would otherwise have expired; the agents are not permitted to disagree hence the model has no representation of fission events.

Agent-based modelling has also been used to simulate collective nest site selection in eusocial insects such as bees [List et al., 2009] and ants [Pratt et al., 2005]. A significant differentiator between our work and previous applications of agent-based models to the study of decision making in natural groups is that we have not used simulation. Rather, our methodology has been to use formal verification.

Whilst, arguably, simulation is far less computationally expensive than model checking the latter has one distinct advantage: completeness. It is rarely practical to explore the entirety of the model parameter space using simulation. Even where exhaustive enumeration can be achieved this in no way guarantees completeness since, for non-deterministic systems (such as ours), each parameter set may correspond to a vast number of potential model runs [Girard and Pappas, 2006].

Since model checking is a resource intensive activity we, necessarily, have considered a model with a comparatively small parameter space. However, we have been able to use this to our advantage: since the parameter space is manageable we have been able to exhaustively enumerate all possible instances of our model. Furthermore, by employing formal verification techniques our results are comprehensive. We have explored our model in its entirety.

We have formalised the group location as a temporal consensus action game with goals, T-CAG-G(GLP). In doing so we have shown that the T-CAG-G model can be used to describe real-world problems where agents have diverse preferences over the order in which their goals are achieved, yet the agents achieve these goals whilst remaining in consensus and staying together. Such a setting includes many group decision making scenarios, from choosing activities for a social outing, to addressing international problems such as climate change, or financial instability [Conradt, 2008].

In our formalisation, the quorum function q_t for each state $t \in T$ is specified according to the rationality postulates introduced in section 5.2, and the agents' ordering over their goals. These, in turn, are determined by the drive values, representing the strength of an agent's need to achieve a particular goal. The model parameter d_{max} sets an upper bound for an agent's drives beyond which the agent expires. It is therefore crucial that agents achieve their goals before their needs become too great. In this respect the drives of the agents, combine with the model parameter d_{max} such that each agent has a deadline by which each goal must be achieved.

Strictly speaking, however, in T-CAG-G(GLP) the agents' goals are not 'achieved' in the sense that the goals are never removed from an agent's goal set. Instead, the most recently realised goal becomes the least urgent in the agents' ordering over goals. Therefore in T-CAG-G(GLP) we model maintenance goals rather than achievement goals. This is consistent with the biological pedigree of the group location problem. For example, when we eat we do not permanently lose the goal to eat, we merely de-prioritise the urgency with which this goal must next be achieved.

Our investigation of T-CAG-G(GLP) has examined the behaviour of the agents under the solution concepts for CAG-Gs (introduced in Chapter 5) and their temporal counter-parts (introduced in Chapter 6). In particular we have focused on strong, weak and q_C-strong consensus paths; although we did not explicitly consider C-weak consensus paths it is the case that, for T-CAG-G(GLP), these are wholly equivalent to weak consensus paths.

We found that consensus paths exist for all of our solution concepts across a broad range of conditions. Of the solution concepts we have examined we found that the agents are most likely to be successful where the consensus action space contains only q_C-minimal strong consensus actions. In particular we found that this refinement of the strong consensus action space not only produces a set of joint actions about which the agents agree more easily (i.e., in fewer rounds), but also leads to more robust decision outcomes and notably increases the chances of the agents' success. This holds across a high proportion of specifications for the quorum functions irrespective of the initial state of the agents.

For each instance of T-CAG-G(GLP) and each solution concept we computed the properties $p(minXca)$ and $p(avgXca)$. Respectively, these are the minimum and average probabilities that from a reachable non-deadlocked state, the subsequent state will also be a non-deadlocked state. Recall that these measures are indicative of the tightness of the consensus action space, hence, how challenging it is for the agents to follow consensus paths.

We found that weak consensus paths, having the lowest means for both $p(minXca)$ and $p(avgXca)$, are the hardest for the agents to follow (see table 7.3). Over all instances, for weak consensus the mean $\overline{p(avgXca)} = 0.7449$ suggesting that on average approximately one in four ($(1 - 0.7449)^{-1} \approx 4$) weak consensus actions lead to a state where further consensus action is not possible. The result is similar for strong consensus where, over all instances, $\overline{p(avgXca)} = 0.7773$. Considering all instances for q_C-strong consensus paths we have $\overline{p(avgXca)} = 0.8769$. This corresponds to only one in eight q_C-minimal strong consensus actions leading to a state where further consensus action is not possible. Thus, in comparison to strong/weak consensus paths, on average over all instances of T-CAG-G(GLP) q_C-strong consensus paths are approximately twice as easy to follow.

A different, yet consistent view arises from considering only those cases where for a given solution concept a consensus path exists. Where a weak consensus path exists the mean $\overline{p(avgXca)} = 0.8105$, where a strong consensus exists $\overline{p(avgXca)} = 0.8720$ and where a q_C-strong consensus path exists $\overline{p(avgXca)} = 0.9866$. These results correspond, respectively and approximately, to one in five, one in eight and one in 75 consensus actions leading to a state where further consensus action is not possible. For T-CAG-G(GLP), where consensus paths exist it is on average approximately nine times easier for the agents to follow a q_C-strong consensus path that it is to follow a strong consensus path. Similarly, it is approximately 15 times easier for the agents to follow a q_C-strong consensus path than it is to follow a weak consensus path.

Considering the property $p(minXca)$ we can also view the tightness of the consensus action space from the worst case perspective. From table 7.3 we can see that over all instances the mean values for $\overline{p(minXca)}$ correspond to approximately one in two consensus actions leading to deadlock in the successor state for both strong and weak consensus; whereas approximately one in seven q_C-minimal consensus actions will lead to a deadlocked state.

Considering values for $\overline{p(minXca)}$ only in cases where consensus paths exist emphasises the ease with which the agents can follow q_C-strong consensus paths. We find that in the worst case one in two weak consensus actions lead to deadlock states. Strong consensus is slightly more robust; at worst, one in three strong consensus actions lead to a deadlocked state. However, for the q_C-minimal strong consensus action space only one in 25 consensus actions lead to deadlock in the successor state.

Our results show that since weak consensus can be reached in wider array of circumstances than, for example, strong consensus, weak consensus paths exist where strong consensus paths do not. Although, typically, the permissive nature of weak consensus leads to consensus paths that are challenging to follow (since many possible deviations exist) we did find that, on occasions, there are advantages to this modality of consensus. In 348 instances, or just under 1% of all instances the weak consensus action space was tight and no strong (or q_C-strong) consensus path exists.

We considered the effects of the initial state upon the success of the agents. Our results support views regarding the effects of synchrony on group decisions as expressed in the biological literature [Conradt and Roper, 2003]. In particular we found that as the synchrony of the agents increases not only are consensus paths more likely to exist (see table 7.5) but they also become easier for the agents to follow (see table 7.6). This is true irrespective of solution concept.

We investigated a representative selection of 36 initial states and categorised these into three initial state kinds corresponding to the agents' drives being fully, partly or not synchronised. For each initial state kind and solution concept we computed the mean properties $\overline{p(minXca)}$ and $\overline{p(avgXca)}$. Table 7.6 shows these data when taken over all instances and only those where a relevant consensus path exists. These results show that in every initial state kind both $\overline{p(minXca)}$ and $\overline{p(avgXca)}$ are lowest for weak consensus and greatest for q_C-strong consensus; values for strong consensus lie between the two. Thus for all initial state kinds it is easiest for the agents to follow q_C-strong consensus paths and hardest for the agents to follow weak consensus paths.

For each solution concept we found that both $\overline{p(minXca)}$ and $\overline{p(avgXca)}$ are at their lowest where the agents are not synchronised in the initial state. The means of both properties increase as the agents' drives become more synchronised in the initial state, reaching their greatest values where the agents' drives are initially fully synchronised. The results show that for all solution concepts, in both the average and worst case terms, it is easiest for agents to follow consensus paths where their drives are fully synchronised in the initial state. Decreasing the degree of synchrony consistently leads to consensus paths that are harder for the agents to follow.

An interesting question is: on average, how much easier is it for the agents to follow consensus paths from initial states where they are synchronised in comparison to where they are not? Taking our results for $\overline{p(avgXca)}$ where a consensus path exists we find that: for weak consensus, where the agents are not initially synchronised approximately one in five consensus actions lead to a deadlocked state. This decreases to approximately one in six where the agents are fully synchronised in the initial state. For strong consensus, where the agents are not synchronised approximately one in seven consensus actions result in a deadlock state; this decreases to approximately one in ten where the agents are initially synchronised. For q_C-strong consensus paths we find that where the agents are not initially synchronised approximately one in 71 consensus actions will lead to deadlocked states; where the agents are fully synchronised in the initial state this decreases to approximately one in 93. To place these results in context it is useful to recall that even from a state where the agents are initially not synchronised it is necessary that the agents perform only three consensus actions (at minimum) in order that their drives become fully synchronised.

Interestingly, our results suggest that consensus paths are least likely to exist from initial states where the agents' drive values give rise to orderings that are rotations of each other. In the two initial states, #8 and #13, where consensus paths (in any solution concept) are least likely to exist, the agents' orderings over their goals form Condorcet cycles [Condorcet, 1785]. Table 7.10, below, shows the relevant portions of the agents' orderings over goals for these initial states, note that these states are symmetrical about agents 2 and 3.

Table 7.10: Initial states where the agents' orderings form Condorcet cycles.

Initial state #	$\succeq_i, i = 1$	$\succeq_i, i = 2$	$\succeq_i, i = 3$
8	$g_3 > g_2 > g_1$	$g_1 > g_3 > g_2$	$g_2 > g_1 > g_3$
13	$g_3 > g_2 > g_1$	$g_2 > g_1 > g_3$	$g_1 > g_3 > g_2$

The cyclical nature of the agents' orderings present what has become known in the social choice community as a voting paradox. Consider initial state #8: agents 1 & 2 both agree that $g_3 > g_2$ but agent 3 does not. Agents 2 & 3 agree that $g_1 > g_3$ but agent

1 does not; similarly, agents 1 & 3 agree that $g_2 > g_1$ whilst agent 2 does not. For orderings of this kind the, commonly employed, simple majority voting rule cannot determine a social ordering or an overall winner; in other words the rule is *indecisive* [Dasgupta and Maskin, 2008].

An analogue for decisiveness in our model is found where the consensus action space is tight. Although such a condition is not wholly decisive in that for a given state more than one consensus action may exist, where the consensus action space is tight *all* consensus actions share the property that they exist on a consensus path. Thus the agents are guaranteed to be successful. Interesting, even for these troublesome initial states the consensus action space can be tight; and this is true for all solution concepts. In particular a q_C-strong consensus path exists from these initial states for 80% of group strategies. Surprisingly, we find that in over 90% of these cases the q_C-minimal strong consensus action space is tight.

We looked at 1,000 group strategies, and so exhaustively explored the space of rational quorum functions for T-CAG-G(GLP). We found that a large proportion of group strategies lead to the existence of consensus paths irrespective of initial state. We found that the q_C-strong consensus path solution concept is particularly resilient to both the specification of the quorum function and the choice of initial state, see table 7.8. For this solution concept the consensus action space was tight, irrespective of initial state, for approximately four (in the case of strong consensus) to six (in the case of weak consensus) times as many strategies.

Individually, each agent, $i \in A$ follows one of ten strategies (denoted σ_i) describing the specification of their quorum values (see table 7.1). Where $\sigma_i = 0$ the agent unconditionally supports all joint actions, such agents behave in a selfless manner, ceding to the will of the other agents. In contrast where $\sigma_i = 9$ we can say that agent i is selfish, unconditionally supporting only the joint action that achieves its non-dominated goal and objecting to all others. Broadly speaking, as σ_i increases the agent becomes increasingly independent.

Independence and interdependence in collective decision making is examined in List et al. [2009]. They also use an agent-based model, however, their interest is in nest-site selection in honeybees. Their model includes a weighting parameter (λ) controlling the extent to which each agent respects the opinions of other agents (regarding the quality of potential nest sites). They found that only in populations where both independent and interdependent behaviours are present do the agents make reliable, accurate decisions.

We too found, for strong and q_C-strong consensus paths, strategies that are neither too interdependent (σ_i is low) or too independent (σ_i is high) produce the best outcomes for the agents (see figure 7.2). In particular we found, for these solution concepts, that given the homogeneous strategies $\sigma_A = \langle 4, 4, 4 \rangle$ or $\sigma_A = \langle 5, 5, 5 \rangle$ not only is a consensus path guaranteed to exist (irrespective of initial state) the consensus action space is also guaranteed to be tight. That is: under no circumstances will these group strategies lead to the agents being unsuccessful. Interestingly, the homogeneous strategy $\sigma_A = \langle 5, 5, 5 \rangle$ corresponds to a strengthening of the postulate of unconditional support such that quorum values decrease with strong (as opposed to weak) monotonicity given strictly (as opposed to non-strictly) increasing rank-order.

The game theoretic model of Conradt and Roper [2007] suggests that, due to evolutionary pressures, consensus (shared) decision making should be far more widespread than dictatorial (unshared) decision making in natural groups. In particular their model predicts that dictatorships only evolve where: the costs of performing an activity too early are similar or lower than the costs of performing an activity too late, the group is relatively homogeneous with respect to activity budgets and the group is not above its optimal size.

These conditions resemble those found in T-CAG-G(GLP). Although we do not explicitly quantify costs and benefits in our model it is clear that the cost of performing an activity too early, e.g., moving to location one where presently location three is most preferred, is significantly less than performing it too late e.g., where an agent's drive equals d_{max} and it expires. In T-CAG-G(GLP) the drives of the agents operate in an identical manner hence the group is homogeneous with respect to activity budgets. Since in T-CAG-G(GLP) there are no costs associated with group size, the group is not above its optimal size. Our results suggest that groups containing a dictator can indeed be successful under these circumstances. Not only do consensus paths exist for all solution concepts for over 75% of dictatorial group strategies we found in every case that the consensus action space was tight.

Lastly, we investigated relationships between the mean quorum value and the likelihood of consensus path existence and tightness of the consensus action space. We confirmed our intuitive hypothesis that a lower mean quorum value increases the likelihood of consensus path existence. We found the mean quorum value to be a significant predictor of a tight consensus action space although we found little evidence suggesting that a lower mean quorum value necessarily leads to this. We believe that these results demonstrate the potential of our models to enable original research into the behaviour of groups where individual behaviours are conditioned upon threshold responses.

Chapter 8

Conclusions

In this thesis we have presented a family of models for describing the behaviour of multi-agent systems in which self-interested, rational agents make group decisions using a biologically inspired consensus mechanism. Each model has addressed a particular aspect of collective behaviour in agents. Our games describe coalition formation (CGs), joint action (CAGs) and goal achievement (CAG-Gs), over time (T-CAG-Gs).

We have reviewed, in Chapter 2, related work from the domains of economics, political science, computer science, sociology and empirical/theoretical biology. We found that whilst collective decision making through threshold mediated behaviours is commonplace in the natural world, and is gaining traction in the technological world, existing models of collective decisions for multi-agent systems are not well suited to describing this modality of consensus decision making. This has been the gap we have tried to address in this thesis.

In Chapter 3 we introduced consensus games (CGs) and studied a problem at the heart of cooperative behaviour in multi-agent systems; that of coalition formation. Following from game-theoretic research we presented two, complementary individually rational solution concepts, strong and weak consensus, and formulated their collectively rational counterparts. Through a series of decision problems we considered the verification, existence and non-existence of outcomes satisfying each solution concept. This characterisation identified aspects of threshold-based decision making which can be computed efficiently, in time that is polynomial in the number of agents. In particular the individually rational solution concepts of strong and weak consensus fall into this category. We also used knowledge-based protocols to show that, for distributed settings, both strong and weak consensus can be established using broadcast messaging. For either solution concept each agent is required to make at most one announcement.

The CAG model presented in Chapter 4 extends the CG model to incorporate consensual joint action by groups of agents. We found that the introduction of joint actions has implications for notions of collective rationality and that certain solution concepts applicable for CGs were not desirable for CAGs. Hence new solution concepts for this, and subsequent models, were introduced. Whilst the inclusion of joint actions in the model had inevitable effects on the computational complexity of various decision problems the problem of verification for strong and weak consensus remained tractable.

We have considered several aspects of rationality in these games. In addition to the individually and collectively rational solution concepts for our games, in Chapter 5 we also considered how rational threshold values can be selected for agents where agents have an ordering over the sets of goals they may achieve. The introduction of rationality postulates placed constraints upon the specification of threshold values for the agents and also affected the computational complexity of certain decision problems. In particular we found that the verification problem for the collectively rational solution concept of the q_C-minimal core in CAG-Gs can be solved in time polynomial in the number of agents.

In Chapter 6 we introduced T-CAG-Gs. These transition systems describe repeated games, played over time. In this model we considered the implications of composite actions - where groups of agents act in consensus with themselves, but not with each other. The model was extended to accommodate such behaviours where the actions agents may perform are constrained and their outcomes, in terms of the goals that they achieve, may be contradictory. We introduced consensus paths, temporal forms for our solution concepts and considered the computational complexity of verification problems for these. We found that T-CAG-G models are well suited to existing model checking techniques and showed that decision problems concerning the existence of consensus paths can be efficiently solved.

Throughout this work the emphasis has been on creating models that are practical and can be applied to real-world situations. In Chapter 7 we demonstrated this by applying the T-CAG-G model to a biologically inspired problem, that of maintaining cohesion in heterogeneous groups. Our results were compared with those from both theoretical and empirical research into collective behaviours in animal groups. We found that the predictions of our model concur with current thinking from those domains. This suggests that our models may be of significant interest and benefit to researchers active in these areas.

8.1 Summary of Contributions

The main contributions of this work have been to:

- define a series of formal models for describing multi-agent systems where agents' interactions are mediated by thresholds,
- propose novel solution concepts for these models capturing notions of both individual and collective rationality,
- provide a characterisation of the computational complexity of decision problems for these solution concepts,
- demonstrate the application of the models to the investigation of consensus, coordination and cooperation in natural groups.

8.2 Future Work

Ours is a first attempt to formalise models and solution concepts for collective action mediated by individual thresholds in multi-agent systems. As such there are many possibilities which either for reasons of time or scope were not addressed here. We briefly consider some of these below.

In Chapter 3 we defined knowledge based protocols for reaching consensus in a distributed setting. These protocols assume that communication between agents is broadcast (global). Other research has considered threshold behaviours where agents' communication is restricted by a network topology, e.g., [Chwe, 1999; Kempe et al., 2003; Watts and Strogatz, 1998]. Such circumstances have been considered only for binary decisions and have not considered aspects such as coalition formation, joint action or goal achievement. It would therefore be interesting to consider variants of our models in which agent interactions are strictly local.

One aspect of collective decision making that is of great interest to those concerned with computational social choice is strategic manipulation through misreporting of preference. In threshold models the obvious analogous practice is the misreporting of threshold values. For our models, strategic manipulations attempt to advantageously increase/reduce the likelihood that some consensus outcome occurs. For temporal games such behaviour could influence consensus paths and hence the set of states which may be reachable.

On the one hand agents in our models have a great deal of power, a single agent can veto a coalition/joint action (changing these from consensus to non-consensus outcomes) simply by objection. On the other hand, in general it is very difficult for an agent to effect a manipulation such that a non-consensus outcome becomes a consensus outcome. This suggests that coalitional manipulations, as in Beckmann et al. [2009], may be of particular interest for our models. The potential for strategic manipulation in CAG-Gs is largely dependent upon the specifications of joint actions and the goals these achieve. It may be interesting therefore to investigate specifications for these structures that are immune to strategic manipulations.

Whilst we have examined how threshold values may be rationally selected for the agents, our notions of rationality are rather weak. As we have shown in Chapter 7 a wide range of threshold strategies from extreme inter-dependence (unconditionally support everything) to extreme independence (only support the non-dominated outcome) coincide with our rationality postulates. An interesting result from Chapter 7 is that, for the problem we studied, one of our rationality postulates can be strengthened.

The postulate of proportional support requires that agents' quorum values increase with weak monotonicity as their keenness to participate in joint actions decreases. For both strong and q_C-strong consensus paths we found that where this response is strongly monotonic the agents are guaranteed to be successful, regardless of their initial state. This result raises several natural questions regarding the use of this stronger postulate in wider circumstances. Most obviously, we might investigate whether this result holds across a broad selection of the parameter space for the T-CAG-G(GLP) problem. More generally it would also be interesting to investigate the characteristics of problems where a *strongly-proportional* strategy does/does not lead to the best possible outcomes for the agents.

Our focus has been on models where agents' thresholds are specified by a deterministic quorum function. However for many natural systems the quorum response is described as a stochastic process. In these systems the *probability* of an individual supporting some collective act is proportional to the number others also showing support. One possible extension of our work would to consider such probabilistic quorum responses. It would also be interesting to consider situations where agents' thresholds may be adaptive and so be selected, for example, based on previous experiences. Such an approach has already been applied to minority games [Challet and Zhang, 1997].

We have demonstrated that our models are useful for applied research. Our work has considered the group location problem in a particularly abstract and simplified form. In future developments of this strand of applied work it would be interesting to increase the realism of both the model of the environment and the agents' drives, as for example in the work of Sellers et al. [2007].

In particular, our model of the environment treats all locations equally. In reality natural groups inhabit diverse environments where locations are, for example, not equidistant from each other, therefore certain locations may require significant and repeated commitment from the group in order that they are reached. Our current model does not capture this. Similarly, our model for the agents' drives is also a extremely simple. We have assumed that needs grow linearly and at the same rate for all agents and for all goals. It is unlikely that for naturally occurring groups this would be the case.

Increasing the realism of the model also suggests that future work should consider groups of a more realistic group size. This may prove problematic for our current experimental methodology since the size of the state space for the problem is exponential in (amongst other parameters) the number of agents modelled. For groups containing a large number of individuals it is unlikely that the resultant model checking problem would be tractable. To overcome issues of scalability an alternative to our current approach would be to use simulation techniques.

We have explored only one of a number of potential avenues where our work may be applied. However, threshold models have been shown to describe a number of collective behaviours, e.g., [Granovetter, 1978; Granovetter and Soong, 1986; Schelling, 1969]. In future work we look forward to applying our models to domains such as social networking, viral marketing, consumer trend adoption and collaborative recommendation.

8.3 Dissemination

Chapter 3 of this thesis appeared in a modified form as:

Zappala, Julian, Natasha Alechina, and Brian Logan. *Consensus Games* In *Proceedings of the Third Workshop on Cooperative Games in Multiagent Systems (CoopMAS-2012)*, Edited by Stephane Airiau, Yoram Bachrach, Edith Elkind and Lirong Xia. Valencia, Spain, 2012.

and as an extended abstract in:

Zappala, Julian, Natasha Alechina, and Brian Logan. *Consensus Games (Extended Abstract)* In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Edited by Vincent Conitzer, Michael Winikoff, Lin Padgham and Wiebe van der Hoek. Vol. 3. Valencia, Spain: IFAAMAS, 2012.

Chapter 4 of this thesis appeared in a modified form as:

Zappala, Julian, Natasha Alechina, and Brian Logan (2011). *Consensus Action Games*. In *Proceedings of the IJCAI-2011 Workshop on Social Choice and Artificial Intelligence*: IJCAI 2011.

Appendix A

This appendix contains additional information pertaining to the applied work of Chapter 7.

Initial states

Table 1 gives the specification of the agents' drives in each of the initial states. For each agent $i \in A$, where $A = \{1, 2, 3\}$, the table shows the value of agent i's drive, $d_{i,k}$, where $k \in \{1, 2, 3\}$ in the initial state.

In total we considered 36 configurations of the agents' drives in the initial state. Of these, there is one initial state in which the agents' drives fully synchronised (Full), 15 where the drives are partly synchronised (Part) and 20 where the drives are not synchronised (None). For ease of reference we assign each initial state a number (shown in the column #)

Table 1: List of initial states.

	$i = 1$			$i = 2$			$i = 3$			
#	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	Sync
0	0	1	2	2	1	0	2	1	0	Part
1	0	1	2	1	2	0	2	1	0	None
2	0	1	2	2	0	1	2	1	0	None
3	0	1	2	0	2	1	2	1	0	None
4	0	1	2	1	0	2	2	1	0	None
5	0	1	2	0	1	2	2	1	0	Part
6	0	1	2	2	1	0	1	2	0	None

continued on next page...

Table 1 – continued from previous page

7	0	1	2	1	2	0	1	2	0	Part
8	0	1	2	2	0	1	1	2	0	None
9	0	1	2	0	2	1	1	2	0	None
10	0	1	2	1	0	2	1	2	0	None
11	0	1	2	0	1	2	1	2	0	Part
12	0	1	2	2	1	0	2	0	1	None
13	0	1	2	1	2	0	2	0	1	None
14	0	1	2	2	0	1	2	0	1	Part
15	0	1	2	0	2	1	2	0	1	None
16	0	1	2	1	0	2	2	0	1	None
17	0	1	2	0	1	2	2	0	1	Part
18	0	1	2	2	1	0	0	2	1	None
19	0	1	2	1	2	0	0	2	1	None
20	0	1	2	2	0	1	0	2	1	None
21	0	1	2	0	2	1	0	2	1	Part
22	0	1	2	1	0	2	0	2	1	None
23	0	1	2	0	1	2	0	2	1	Part
24	0	1	2	2	1	0	1	0	2	None
25	0	1	2	1	2	0	1	0	2	None
26	0	1	2	2	0	1	1	0	2	None
27	0	1	2	0	2	1	1	0	2	None
28	0	1	2	1	0	2	1	0	2	Part
29	0	1	2	0	1	2	1	0	2	Part
30	0	1	2	2	1	0	0	1	2	Part
31	0	1	2	1	2	0	0	1	2	Part
32	0	1	2	2	0	1	0	1	2	Part
33	0	1	2	0	2	1	0	1	2	Part
34	0	1	2	1	0	2	0	1	2	Part
35	0	1	2	0	1	2	0	1	2	Full

Bibliography

Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Temporal qualitative coalitional games. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 177–184, New York, NY, USA, 2006a. ACM. ISBN 1-59593-303-4. doi: http://doi.acm.org/10.1145/1160633.1160662. 14, 92

Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. On the logic of coalitional games. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 153–160. ACM, 2006b. 91

Thomas Ågotnes, Wiebe Van Der Hoek, and Michael Wooldridge. Logics for qualitative coalitional games. *Logic Journal of the IGPL*, 17:299–321(23), 2009. doi: doi:10.1093/jigpal/jzp009. URL `http://www.ingentaconnect.com/content/oup/igpl/2009/00000017/00000003/art00004`. 14

N. Alexiou, S. Basagiannis, P. Katsaros, T. Dashpande, and S. Smolka. Formal analysis of the kaminsky dns cache-poisoning attack using probabilistic model checking. In *Proc. IEEE 12th International Symposium on High-Assurance Systems Engineering (HASE'10)*, pages 94–103, 2010. 95

Rajeev Alur and Thomas A Henzinger. Reactive modules. *Formal Methods in System Design*, 15(1):7–48, 1999. 23

Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. *J. ACM*, 49(5):672–713, 2002. ISSN 0004-5411. doi: http://doi.acm.org/10.1145/585265.585270. 91

K.J. Arrow. A difficulty in the concept of social welfare. *The Journal of Political Economy*, 58(4):328–346, 1950. 14, 15

Robert J Aumann. The core of a cooperative game without side payments. *Transactions of the American Mathematical Society*, pages 539–552, 1961. 12

Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT Press, 2008. ISBN 978-0-262-02649-9. 88

B.E. Beckmann, P.K. McKinley, and D.B. Knoester. Effects of communication impairments on quorum sensing. In *Proceedings of 3rd International Conference on Self-Adaptive and Self-Organizing Systems*, 2009. 22, 132

Martin Beer, Mark D'inverno, Michael Luck, Nick Jennings, Chris Preist, and Michael Schroeder. Negotiation in multi-agent systems. *The Knowledge Engineering Review*, 14(03):285–289, 1999. 11

R.D. Beer, R.D. Quinn, H.J. Chiel, and R.E. Ritzmann. Biologically inspired approaches to robotics: What can we learn from insects? 1997. 2

F. Bernardini, M. Gheorghe, and N. Krasnogor. Quorum sensing P systems. *Theoretical Computer Science*, 371(1-2):20–33, 2007. 21

E. Bonabeau, A. Sobkowski, G. Theraulaz, and J.L. Deneubourg. Adaptive task allocation inspired by a model of division of labor in social insects. *Bio Computation and Emergent Computing*, pages 36–45, 1997. 9

J. Bredin, R.T. Maheswaran, C. Imer, T. Başar, D. Kotz, and D. Rus. A game-theoretic formulation of multi-agent resource allocation. In *Proceedings of the fourth international conference on Autonomous agents*, pages 349–356. ACM New York, NY, USA, 2000. 9

Randal E Bryant. Graph-based algorithms for boolean function manipulation. *Computers, IEEE Transactions on*, 100(8):677–691, 1986. 23

Jerry R Burch, Edmund M Clarke, Kenneth L McMillan, David L Dill, and Lain-Jinn Hwang. Symbolic model checking: 10¡ sup¿ 20¡/sup¿ states and beyond. *Information and computation*, 98(2):142–170, 1992. 23

D. Challet and Y.C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A: Statistical Mechanics and its Applications*, 246(3):407–418, 1997. 21, 132

X. Chen and X. Deng. Settling the complexity of two-player Nash equilibrium. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 261–272. IEEE, 2006. 46

Bailiang Cheng, Guosun Zeng, and Anquan Jie. Adding branching temporal dimension to qualitative coalitional games with preference. *Journal of Computers*, 5(5):749–756, 2010. 14

Y. Chevaleyre, P. E. Dunne, U. Endriss, J. Lang, M. Lemaître, N. Maudet, J. Padget, S. Phelps, J. A. Rodríguez-Aguilar, and P. Sousa. Issues in Multiagent Resource Allocation. *Informatica*, 30:3–31, 2006. 9

Michael Suk-Young Chwe. Structure and strategy in collective action 1. *American Journal of Sociology*, 105(1):128–156, 1999. 17, 44, 45, 131

Edmund M. Clarke, E Allen Emerson, and A Prasad Sistla. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 8(2):244–263, 1986. 23

Edmund M Clarke, Orna Grumberg, and David E Long. Model checking and abstraction. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 16 (5):1512–1542, 1994. 23

PR Cohen and HJ Levesque. Teamwork. *Noûs(Bloomington, Indiana)*, 25(4):487–512, 1991. 16, 62

Marie Jean Antoine Nicolas Caritat Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale, 1785. 125

V. Conitzer and T. Sandholm. Nonexistence of voting rules that are usually hard to manipulate. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 627. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006. 15

L. Conradt. Group decisions: How (not) to choose a restaurant with friends. *Current Biology*, 18(24):1139–1140, 2008. 122

L. Conradt and C. List. Introduction. group decisions in humans and animals: a survey. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364:719–742, 2009. 17

L. Conradt and TJ Roper. Group decision-making in animals. *Nature*, 421(6919): 155–158, 2003. doi: doi:10.1038/nature01294. 94, 124

L. Conradt and T.J. Roper. Democracy in animals: the evolution of shared group decisions. *Proceedings of the Royal Society B: Biological Sciences*, 274(1623): 2317–2326, September 2007. URL `http://dx.doi.org/10.1098/rspb.2007.0186`. 94, 127

L. Conradt and T.J. Roper. Conflicts of interest and the evolution of decision sharing. *Philosophical Transactions B*, 364(1518):807, 2009. 94

Larissa Conradt and Timothy J. Roper. Consensus decision making in animals. *Trends in Ecology & Evolution*, 20(8):449–456, August 2005. URL `http://www.sciencedirect.com/science/article/B6VJ1-4G9GMT7-1/2/fbb4083cc9380bc778bdbdc20b625c23`. 94

James C Corbett, Matthew B Dwyer, John Hatcliff, Shawn Laubach, Corina S Pasareanu, Hongjun Zheng, et al. Bandera: Extracting finite-state models from java source code. In *Software Engineering, 2000. Proceedings of the 2000 International Conference on*, pages 439–448. IEEE, 2000. 23

Partha Dasgupta and Eric Maskin. On the robustness of majority rule. *Journal of the European Economic Association*, 6(5):949–973, 2008. 126

C. Daskalakis and C. Papadimitriou. Computing equilibria in anonymous games. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 83–93. IEEE, 2007. 21, 45

Simon DeDeo, David C Krakauer, and Jessica C Flack. Inductive game theory and the dynamics of animal conflict. *PLoS computational biology*, 6(5):e1000782, 2010. 94

Maria Pamela Dobay, Akos Dobay, Johnrob Bantang, and Eduardo Mendoza. How many trimers? modeling influenza virus fusion yields a minimum aggregate size of six trimers, three of which are fusogenic. *Molecular BioSystems*, 7(10):2741–2749, 2011. 95

K. Dowding and M. van Hees. In Praise of Manipulation. *Science*, 21:769–803, 1977. 15

A.J. Dowell, M. Wooldridge, and P. McBurney. The Computational Difficulty of Bribery in Qualitative Coalitional Games. *Fondazione Eni Enrico Mattei Working Papers*, page 159, 2007. 14

M. Duflot, M. Kwiatkowska, G. Norman, and D. Parker. A formal analysis of Bluetooth device discovery. *Int. Journal on Software Tools for Technology Transfer*, 8 (6):621–632, 2006. 95

P.E. Dunne and M. Wooldridge. Preferences in qualitative coalitional games. In *In Proceedings of the Sixth Workshop on Game Theoretic and Decision Theoretic Agents (GTDT-04)*, volume 4, pages 29–38, New York, NY, July 2004. 14, 78

P.E. Dunne, W. van der Hoek, S. Kraus, and M. Wooldridge. Cooperative boolean games. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 2*, pages 1015–1022. International Foundation for Autonomous Agents and Multiagent Systems, 2008. 13

Emad Eldeen Elakehal and Julian Padget. Msmas: Modelling self-managing multi agent systems. *Scalable Computing: Practice and Experience*, 13(2), 2012. 10

Edith Elkind and Michael Wooldridge. Hedonic coalition nets. In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pages 417–424. International Foundation for Autonomous Agents and Multiagent Systems, 2009. 13

E Allen Emerson and Edmund M Clarke. Using branching time temporal logic to synthesize synchronization skeletons. *Science of Computer programming*, 2(3):241–266, 1982. 88

R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Mass., 1995. 38

A. Fehnker and P. Gao. Formal verification and simulation for performance analysis for probabilistic broadcast protocols. In *Proc. 5th International Conference on Ad-Hoc, Mobile, and Wireless Networks (ADHOC-NOW'06)*, volume 4104 of *LNCS*, pages 128–141. Springer, 2006. 95

Steven A Frank. Evolutionary foundations of cooperation and group cohesion. In *Games, Groups, and the Global Good*, pages 3–40. Springer, 2009. 94

S. Franklin and A. Graesser. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *Proceedings of the Third International Workshop on Agent Theories, Architectures and Languages*, volume 1193, pages 21–35. Springer, 1996. 1

N.R. Franks, E.B. Mallon, H.E. Bray, M.J. Hamilton, and T.C. Mischler. Strategies for choosing between alternatives with different attributes: exemplified by house-hunting ants. *Animal Behaviour*, 65(1):215–223, 2003. 19

C. Fuqua, S.C. Winans, and E.P. Greenberg. Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annual Reviews in Microbiology*, 50(1):727–751, 1996. 18

A. Gibbard. Manipulation of voting schemes: a general result. *Econometrica: Journal of the Econometric Society*, pages 587–601, 1973. 15

Antoine Girard and George J Pappas. Verification using simulation. In *Hybrid Systems: Computation and Control*, pages 272–286. Springer, 2006. 121

M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978. 17, 25, 44, 45, 133

M. Granovetter and R. Soong. Threshold models of interpersonal effects in consumer demand. *Journal of Economic Behavior & Organization*, 7(1):83–99, 1986. 17, 133

J. Greensmith and U. Aickelin. The deterministic dendritic cell algorithm. In *7th International Conference on Artificial Immune Systems (ICARIS)*, pages 291–303. Springer, 2008. 2

BJ Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial intelligence*, 86(2):269–357, 1996. 16, 62

B.J. Grosz and S. Kraus. The evolution of SharedPlans. *Foundations of Rational Agency*, 14:227–262, 1999. 15, 16, 62

B.J. Grosz and C.L. Sidner. Plans for discourse. *Intentions in Communication*, pages 417–444, 1990. 1, 16, 62

Hans Hansson and Bengt Jonsson. A logic for reasoning about time and reliability. *Formal aspects of computing*, 6(5):512–535, 1994. 24

J.C. Harsanyi. On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution*, pages 179–196, 1961. 79, 80

John C Harsanyi. Games with incomplete information played by bayesian players, i–iii: Part i. the basic model&. *Management science*, 50(12 supplement):1804–1817, 2004. 46, 47

Klaus Havelund and Thomas Pressburger. Model checking java programs using java pathfinder. *International Journal on Software Tools for Technology Transfer*, 2(4): 366–381, 2000. 23

Klaus Havelund, Mike Lowry, and John Penix. Formal analysis of a space-craft controller using spin. *Software Engineering, IEEE Transactions on*, 27(8):749–765, 2001. 23

J. Heath, M. Kwiatkowska, G. Norman, D. Parker, and O. Tymchyshyn. Probabilistic model checking of complex biological pathways. *Theoretical Computer Science*, 319(3):239–257, 2008. 95

Thomas Herault, Richard Lassaigne, and Sylvain Peyronnet. Apmc 3.0: Approximate verification of discrete and continuous time markov chains. In *Quantitative Evaluation of Systems, 2006. QEST 2006. Third International Conference on*, pages 129–130. IEEE, 2006. 24

RA Hill and PC Lee. Predation risk as an influence on group size in cercopithecoid primates: implications for social structure. *Journal of Zoology*, 245(04):447–456, 1998. 93

Russell A. Hill, Brian Logan, William I. Sellers, and Julian Zappala. An agent-based model of group decision making in baboons. In Anil K. Seth, Tony J. Prescott, and Joanna J. Bryson, editors, *Modelling Natural Action Selection*, pages 454–476. Cambridge University Press, 2011. ISBN 9781107000490. 20

Gerard J Holzmann. The model checker spin. *Software Engineering, IEEE Transactions on*, 23(5):279–295, 1997. 23

L. Hunsberger and M. Zancanaro. A mechanism for group decision making in collaborative activity. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, pages 30–35. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2000. 16

E.B. Jacob, I. Becker, Y. Shapira, and H. Levine. Bacterial linguistic communication and social intelligence. *TRENDS in Microbiology*, 12(8):366–372, 2004. 2

Joost-Pieter Katoen, Ivan S Zapreev, Ernst Moritz Hahn, Holger Hermanns, and David N Jansen. The ins and outs of the probabilistic model checker mrmc. *Performance evaluation*, 68(2):90–104, 2011. 24

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003. 22, 131

J. Kennedy, R.C. Eberhart, et al. Particle swarm optimization. In *Proceedings of IEEE international conference on neural networks*, volume 4, pages 1942–1948. Piscataway, NJ: IEEE, 1995. 2

Andrew J. King, Caitlin M.S. Douglas, Elise Huchard, Nick J.B. Isaac, and Guy Cowlishaw. Dominance and affiliation mediate despotism in a social primate. *Current Biology*, 18(23):1833 – 1838, 2008. ISSN 0960-9822. doi: DOI:10.1016/j.cub.2008.10.048. URL `http://www.sciencedirect.com/science/article/B6VRT-4TYRF97-8/2/1af3a586a492cfa5e40377171a0ea6c5`. 94

Matthias Klusch and Andreas Gerber. Dynamic coalition formation among rational agents. *IEEE Intelligent Systems*, 17(3):42–47, 2002. ISSN 1541-1672. doi: http://dx.doi.org/10.1109/MIS.2002.1005630. 9

N. Krasnogor, M. Gheorghe, G. Terrazas, S. Diggle, P. Williams, and M. Camara. An appealing computational mechanism drawn from bacterial quorum sensing. *Bulletin of the European Association of Theoretical Computer Science*, (85):135–148, February 2005. URL `http://www.cs.nott.ac.uk/~nxk/PAPERS/eatcs-qs.pdf`. 21

S. Kraus. Negotiation and cooperation in multi-agent environments. *Artificial Intelligence*, 94(1-2):79–97, 1997. 1

M. Kwiatkowska, G. Norman, and R. Segala. Automated verification of a randomized distributed consensus protocol using Cadence SMV and PRISM. In G. Berry, H. Comon, and A. Finkel, editors, *Proc. 13th International Conference on Computer Aided Verification (CAV'01)*, volume 2102 of *LNCS*, pages 194–206. Springer, 2001. 95

M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011. 24, 95, 105

M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic verification of hermans self-stabilisation algorithm. *Formal Aspects of Computing*, 24(4):661–670, 2012. 95

Jean-Baptiste Leca, Nolle Gunst, Bernard Thierry, and Odile Petit. Distributed leadership in semifree-ranging white-faced capuchin monkeys. *Animal Behaviour*, 66(6):1045–1052, December 2003. URL `http://www.sciencedirect.com/science/article/B6W9W-49J8TBN-P/2/4cfe030ffa2d28bf0087429575998c91`. 93

H. J. Levesque, P. R. Cohen, and J. H. T. Nunes. On acting together. In *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90)*, pages 94–99. Boston, MA, 1990. 16, 62

Pietro Liò, Emanuela Merelli, and Nicola Paoletti. Multiple verification in computational modeling of bone pathologies. In *Proc. 3rd International Workshop on Computational Models for Cell Processes (COMPMOD'11)*, volume 68 of *EPTCS*, pages 82–96, 2011. 95

C. List, C. Elsholtz, and T.D. Seeley. Independence and interdependence in collective decision making: an agent- based model of nest-site choice by honeybee swarms. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 364(1518):755–762, 2009. 19, 121, 126

Christian List. Democracy in animal groups: a political science perspective. *Trends in Ecology & Evolution*, 19(4):168–169, April 2004. URL `http://www.sciencedirect.com/science/article/B6VJ1-4BS0FBG-2/2/1b986b46e3a19e3d02ba44c98452bdaa`. 2

Alessio Lomuscio, Hongyang Qu, and Franco Raimondi. Mcmas: A model checker for the verification of multi-agent systems. In *Computer Aided Verification*, pages 682–688. Springer, 2009. 23

Andrew McLennan and Rabee Tourky. Simple complexity from imitation games. *Games and Economic Behavior*, 68(2):683 – 688, 2010. ISSN 0899-8256. doi: DOI:10.1016/j.geb.2009.10.003. URL http://www.sciencedirect.com/science/article/B6WFW-4XF83TF-2/2/fc72fad5d003de68b5b9ea6f8dffc539. 20, 45

Nimrod Megiddo and Christos H Papadimitriou. On total functions, existence theorems and computational complexity. *Theoretical Computer Science*, 81(2):317–324, 1991. 46

D. Monderer and L.S. Shapley. Potential games. *Games and economic behavior*, 14: 124–143, 1996. 21

J. Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951. 11

U. Ndukwu and A. McIver. An expectation transformer approach to predicate abstraction and data independence for probabilistic programs. In *Proc. 8th Workshop on Quantitative Aspects of Programming Languages (QAPL'10)*, 2010. 95

H. Nurmi. *Voting procedures under uncertainty*. Springer Verlag, 2002. 15

Hyacinth S Nwana, Lyndon C Lee, and Nicholas R Jennings. Coordination in software agent systems. *British Telecom Technical Journal*, 14(4):79–88, 1996. 10

C. H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994. 56

C.H. Papadimitriou and D. Wolfe. The complexity of facets resolved. In *Foundations of Computer Science, 1985., 26th Annual Symposium on*, pages 74–78. IEEE, 1985. 61

Marc Pauly. A modal logic for coalitional power in games. *Journal of logic and computation*, 12(1):149–166, 2002. 91

Maxim Peysakhov, Christopher Dugan, Pragnesh Jay Modi, and William Regli. Quorum sensing on mobile ad-hoc networks. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 1104–1106, New York, NY, USA, 2006. ACM. ISBN 1-59593-303-4. doi: http://doi.acm.org/10.1145/1160633.1160831. 2, 21

MD Peysakhov and WC Regli. Ant inspired server population management in a service based computing environment. In *Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE*, pages 357–364, 2005. 21

G. Pigozzi. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006. 1

Amir Pnueli. The temporal logic of programs. In *Foundations of Computer Science, 1977., 18th Annual Symposium on*, pages 46–57. IEEE, 1977. 23

S. C. Pratt. Quorum sensing by encounter rates in the ant Temnothorax albipennis. *Behavioral Ecology*, 16(2):488, 2005. 19

S. C. Pratt and D. J. T. Sumpter. A tunable algorithm for collective decision-making. *Proceedings of the National Academy of Sciences*, 103(43):15906, 2006. URL `http://www.pnas.org/content/103/43/15906.abstract`. 2, 19

S. C. Pratt, D. J. T. Sumpter, E. B. Mallon, and N. R. Franks. An agent-based model of collective nest choice by the ant temnothorax albipennis. *Animal Behaviour*, 70 (5):1023–1036, 2005. 121

S.A. Rands, G. Cowlishaw, R.A. Pettifor, J.M. Rowcliffe, and R.A. Johnstone. The emergence of leaders and followers in foraging pairs when the qualities of individuals differ. *BMC Evolutionary Biology*, 8(1):51, 2008. 94

Stephan G. Reebs. Can a minority of informed leaders determine the foraging movements of a fish shoal? *Animal Behaviour*, 59(2):403–409, February 2000. URL `http://www.sciencedirect.com/science/article/B6W9W-45FK6NR-6H/2/fb4c17d4f372e7b44dbc340f5a1f9f5b`. 93

Jeffrey S Rosenschein and Gilad Zlotkin. *Rules of encounter: designing conventions for automated negotiation among computers*. the MIT Press, 1994. 11

R.W. Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973. 21

T.W. Sandholm. Contract types for satisficing task allocation: I theoretical results. In *In AAAI Spring Symposium Series: Satisficing Models*, 1998. 9

M.A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions* 1. *Journal of Economic Theory*, 10(2):187–217, 1975. 15

T .C. Schelling. Models of segregation. *The American Economic Review*, 59(2):488–493, 1969. 2, 17, 133

J.R. Searle. Collective intentions and actions. *Intentions in communication*, pages 401–415, 1990. 16

T.D. Seeley and S.C. Buhrman. Group decision making in swarms of honey bees. *Behavioral Ecology and Sociobiology*, 45(1):19–31, 1999. doi: 10.1007/s002650050536. 18

T.D. Seeley and P.K. Visscher. Choosing a home: how the scouts in a honey bee swarm perceive the completion of their group decision making. *Behavioral Ecology and Sociobiology*, 54(5):511–520, 2003. 18

WI Sellers, RA Hill, and BS Logan. An agent-based model of group decision making in baboons. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1699–1710, 2007. doi: doi:10.1098/rstb.2007.2064. 3, 4, 6, 17, 20, 121, 133

O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, 1998. 9, 63

Peer-Olaf Siebers and Uwe Aickelin. *Encyclopedia of Decision Making and Decision Support Technologies, Vol. 2*, chapter Introduction to Multi-Agent Simulation, pages 445–564. Information Science Reference, 2008. 1

Reid G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *Computers, IEEE Transactions on*, 100(12):1104–1113, 1980. 9

G. Steel. Formal analysis of PIN block attacks. *Theoretical Computer Science*, 367 (1-2):257–270, 2006. 95

D. J. Sumpter and S. C. Pratt. Quorum responses and consensus decision making. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 2008. 17, 19

Emmanuel Tadjouddine. Complexity of verifying game equilibria. In Hans-Dieter Burkhard, Gabriela Lindemann, Rineke Verbrugge, and Lszl Varga, editors, *Multi-Agent Systems and Applications V*, volume 4696 of *Lecture Notes in Computer Science*, pages 103–112. Springer Berlin / Heidelberg, 2007. ISBN 978-3-540-75253-0. URL `http://dx.doi.org/10.1007/978-3-540-75254-7_11`. 46

Milind Tambe. Towards flexible teamwork. *CoRR*, cs.AI/9709101, 1997. 16, 62

A. Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, 1955. 27

R. Tuomela and K. Miller. We-intentions. *Philosophical Studies*, 53(3):367–389, 1988. 16

J. Turek and D. Shasha. The many faces of consensus in distributed systems. *IEEE Computer*, 25(6):8–17, 1992. 5, 45

Peter van Emde Boas. *Machine models and simulations*, pages 1–66. MIT Press, Cambridge, MA, USA, 1990. ISBN 0-444-88071-2. URL `http://dl.acm.org/citation.cfm?id=114872.114873`. 30

Javier Vázquez-Salceda, Huib Aldewereld, and Frank Dignum. Implementing norms in multiagent systems. In *Multiagent system technologies*, pages 313–327. Springer, 2004. 10

Ryan Vogt, John Aycock, and Michael J. Jacobson. Quorum sensing and self-stopping worms. In *WORM '07: Proceedings of the 2007 ACM workshop on Recurring malcode*, pages 16–22, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-886-2. doi: http://doi.acm.org/10.1145/1314389.1314394. 2, 22

J. Von Neumann, O. Morgenstern, H.W. Kuhn, and A. Rubinstein. *Theory of games and economic behavior*. Princeton university press Princeton, NJ, 1947. 12

D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. 131

Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002. 22

G. Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. The MIT press, 2000. 1

M. Wooldridge and N.R. Jennings. Intelligent agents: Theory and practice. *Knowledge engineering review*, 10(2):115–152, 1995. 1

M. Wooldridge, T. Ågotnes, P.E. Dunne, and W. Van der Hoek. Logic for automated mechanism design-a progress report. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, volume 22, page 9. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007. 11, 13

Michael Wooldridge. *An introduction to multiagent systems*. Wiley, 2005. 10

Michael Wooldridge and Paul E. Dunne. On the computational complexity of qualitative coalitional games. *Artificial Intelligence*, 158(1):27–73, 2004. ISSN 0004-3702. doi: http://dx.doi.org/10.1016/j.artint.2004.04.002. 9, 35, 45, 46, 61, 92

M. Youssefmir and B. Huberman. Resource contention in multi-agent systems. In *First International Conference on Multi-Agent Systems (ICMAS-95)*, 1995. 1

J. Zappala. Multi-agent Simulation of Group Decision Making in Animals. Master's thesis, School of Computer Science, University of Nottingham, 2008. 17