Modelling Coalitions: ATL + Argumentation

Nils Bulling, Jürgen Dix
Department of Informatics, Clausthal University of Technology, Germany
bulling@in.tu-clausthal.de,
dix@tu-clausthal.de

Carlos I. Chesñevar
Department of Computer Science and Engineering, Universidad Nacional del Sur, Bahía Blanca, Argentina
cic@cs.uns.edu.ar

ABSTRACT

In the last few years, *argumentation frameworks* have been successfully applied to multi agent systems. Recently, argumentation has been used to provide a framework for reasoning about *coalition formation*. At the same time *alternating-time temporal logic* has been used to reason about the behavior and abilities of coalitions of agents. However, **ATL** operators account only for the *existence of successful strategies* of coalitions. They do not consider whether coalitions can be actually formed.

This paper is an attempt to combine both frameworks and to develop a logic through which we can reason at the same time (1) about abilities of coalitions of agents and (2) about the formation of coalitions. We provide a formal extension of **ATL**, $\mathbf{ATL^c}$, in which the actual computation of the coalition is modelled in terms of argumentation semantics. We show that $\mathbf{ATL^c}$'s proof theory can be understood as a natural extension of the model checking procedure used in **ATL**.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal logic, Temporal logic*

General Terms

Theory, Logical foundations

Keywords

multi-agent systems, argumentation, coalition formation, game theory, temporal logic

1. INTRODUCTION AND MOTIVATIONS

In the context of multiagent systems, *argumentation frameworks* [18, 8] have proven to be useful for several purposes, such as *joint deliberation*, *persuasion*, *negotiation*, and *conflict resolution* [21, 19, 20, 16]. In particular, it has been shown recently that argumentation provides a sound setting to model *reasoning about coalition formation* in multi-agent systems [3, 4]. In this approach, conflict and preference relationships among coalitions are used to define which coalitions should be adopted by the agents. This is done according to a particular argumentation semantics, which can then be computed using a suitable proof theory.

Alternating-time temporal logic (**ATL**) [2] is a temporal logic that is used for reasoning about the behavior and abilities of agents under various rationality assumptions [6, 13, 14]. The key construct in **ATL** has the form $\langle\langle A\rangle\rangle\phi$, which expresses that a coalition A of agents can *enforce* the formula ϕ. Under a model theoretic viewpoint, $\langle\langle A\rangle\rangle\phi$ holds whenever the agents in A have a winning strategy for ensuring the satisfiability of ϕ, independently of the behavior of A's opponents. However, this operator accounts only for the *theoretical existence* of such a strategy as it does not take into account whether the coalition A can be actually formed. Indeed, in order to join a coalition, agents usually require some kind of *incentive* (i.e. sharing common goals or getting rewards), since forming a coalition does not come for free (fees have to be paid, communication costs occur, etc.). Several possible coalition structures among agents may arise, from which the best ones should be adopted according to some rationally justifiable procedure.

This paper presents a first approach towards extending **ATL** for modelling coalitions through argumentation. We provide a formal extension of **ATL**, $\mathbf{ATL^c}$, by including a new construct $\langle\!| A |\!\rangle\phi$ which denotes that *the group A of agents is able to build a coalition B, $A \subseteq B$, such that B can enforce ϕ.* That is, it is assumed that agents in A work together, inviting other agents to join and form a coalition B. This intuition is in accordance with **ATL** where larger coalitions are more powerful than smaller ones. The actual computation of the coalition is modelled in terms of a given argumentation semantics [10] in the context of coalition formation [3]. We show that the proof theory for modelling coalitions in our framework can be embedded as a natural extension of the model checking procedure used in **ATL**.

The rest of the paper is structured as follows. Section 2 summarizes the main concepts of alternating-time temporal logic (**ATL**). In Section 3 we introduce the notion of *coalitional framework* [3] as well as some fundamental concepts from argumentation theory. Section 4 provides an argumentation-based view of coalition formation by merging **ATL** and the coalitional framework introduced in Sections 2 and 3. In Section 5 we turn to model checking and in Section 6 we give an outlook to other alternative semantics. Finally, we discuss related work and conclude.

Cite as: Modelling Coalitions: ATL + Argumentation, Nils Bulling, Carlos I. Chesñevar, Jürgen Dix, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal,pp. 681-688.

2. ATL

Alternating-time temporal logic (**ATL**) [2] enables reasoning about temporal properties and strategic abilities of agents. The language of **ATL** is defined as follows.

Definition 1 ($\mathcal{L}_{ATL}$ [2]). *Let $\mathbb{A}\text{gt} = \{a_1, \ldots, a_k\}$ be a nonempty finite set of all agents, and Π be a set of propositions (with typical element p). We denote by "a" a typical agent, and by "A" a typical group of agents from $\mathbb{A}\text{gt}$. $\mathcal{L}_{ATL}(\mathbb{A}\text{gt}, \Pi)$ is defined by the following grammar: $\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle A\rangle\rangle \bigcirc \varphi \mid \langle\langle A\rangle\rangle \Box\, \varphi \mid \langle\langle A\rangle\rangle \varphi\, \mathcal{U}\, \varphi$.*

Informally, $\langle\langle A\rangle\rangle\varphi$ expresses that agents A have a *collective strategy to enforce* φ. **ATL** formulae include the usual temporal operators: $\bigcirc$ ("in the next state"), $\Box$ ("always from now on") and $\mathcal{U}$ (strict "until"). Additionally, $\Diamond$ ("now or sometime in the future") can be defined as $\Diamond\, \varphi \equiv \top\, \mathcal{U}\, \varphi$.

The semantics of **ATL** is defined by *concurrent game structures.*

Definition 2 (**CGS** [2]). *A* concurrent game structure (**CGS**) *is a tuple $\mathcal{M} = \langle \mathbb{A}\text{gt}, Q, \Pi, \pi, Act, d, o\rangle$, consisting of: a set $\mathbb{A}\text{gt} = \{a_1, \ldots, a_k\}$ of* agents*; set Q of* states*; set Π of* atomic propositions*;* valuation of propositions *$\pi : Q \to \mathcal{P}(\Pi)$; set Act of* actions*. Function $d : \mathbb{A}\text{gt} \times Q \to \mathcal{P}(Act)$ indicates the actions available to agent $a \in \mathbb{A}\text{gt}$ in state $q \in Q$. We often write $d_a(q)$ instead of $d(a, q)$, and use $d(q)$ to denote the set $d_{a_1}(q) \times \cdots \times d_{a_k}(q)$ of* action profiles *in state q. Finally, o is a* transition function *which maps each state $q \in Q$ and action profile $\vec{\alpha} = \langle \alpha_1, \ldots, \alpha_k\rangle \in d(q)$ to another state $q' = o(q, \vec{\alpha})$.*

A *path* $\lambda = q_0 q_1 \cdots \in Q^+$ is an infinite sequence of states such that there is a transition between each q_i, q_{i+1}.We define $\lambda[i] = q_i$ to denote the i-th state of λ. The set of all paths starting in q is defined by $\Lambda_{\mathcal{M}}(q)$.

A (memoryless) *strategy* of agent a is a function $s_a : Q \to Act$ such that $s_a(q) \in d_a(q)$ We denote the set of such functions by Σ_a. A *collective strategy* s_A for team $A \subseteq \mathbb{A}\text{gt}$ specifies an individual strategy for each agent $a \in A$; the set of A's collective strategies is given by $\Sigma_A = \prod_{a \in A} \Sigma_a$ and $\Sigma := \Sigma_{\mathbb{A}\text{gt}}$.

The *outcome* of strategy s_A in state q is defined as the set of all paths that may result from executing s_A: $out(q, s_A) = \{\lambda \in \Lambda_{\mathcal{M}}(q) \mid \forall i \in \mathbb{N}_0\ \exists \vec{\alpha} = \langle \alpha_1, \ldots, \alpha_k\rangle \in d(\lambda[i])\ \forall a \in A\ (\alpha_a = s^a_A(\lambda[i]) \wedge o(\lambda[i], \vec{\alpha}) = \lambda[i+1])\}$, where s^a_A denotes agent a's part of the collective strategy s_A.

The semantics of **ATL** is as follows.

Definition 3 (**ATL** Semantics). *Let a **CGS** $\mathcal{M} = \langle \mathbb{A}\text{gt}, Q, \Pi, \pi, Act, d, o\rangle$ and $q \in Q$ be given. The semantics is given by a satisfaction relation $\models$ as follows:*

$\mathcal{M}, q \models p$ *iff* $p \in \pi(q)$

$\mathcal{M}, q \models \neg\varphi$ *iff* $\mathcal{M}, q \not\models \varphi$

$\mathcal{M}, q \models \varphi \wedge \psi$ *iff* $\mathcal{M}, q \models \varphi$ *and* $\mathcal{M}, q \models \psi$

$\mathcal{M}, q \models \langle\langle A\rangle\rangle \bigcirc \varphi$ *iff there is $s_A \in \Sigma_A$ such that $\mathcal{M}, \lambda[1] \models \varphi$ for all $\lambda \in out(q, s_A)$*

$\mathcal{M}, q \models \langle\langle A\rangle\rangle \Box\, \varphi$ *iff there is s_A such that $\mathcal{M}, \lambda[i] \models \varphi$ for all $\lambda \in out(q, s_A)$ and $i \in \mathbb{N}_0$*

$\mathcal{M}, q \models \langle\langle A\rangle\rangle \varphi\, \mathcal{U}\, \psi$ *iff there is $s_A \in \Sigma_A$ such that, for all $\lambda \in out(q, s_A)$, there is $i \in \mathbb{N}_0$ with $\mathcal{M}, \lambda[i] \models \psi$, and $\mathcal{M}, \lambda[j] \models \varphi$ for all $0 \leq j < i$.*

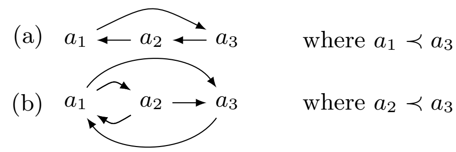

Figure 1: Figure (a) (resp. (b)) corresponds to the coalitional frameworks defined in Example 1 (resp. 3 (b)). Nodes represent agents and arrows between nodes stand for the attack relation.

3. COALITIONS AND ARGUMENTATION

In this section we provide an argument-based characterization of coalition formation that will be used later to extend **ATL**. We follow an approach similar to [3], where an argumentation framework for generating coalition structures is defined. Our approach is a generalization of the framework of Dung for argumentation [10], extended with a *preference relation.* The basic notion is that of a *coalitional framework*, which contains a set of elements $\mathfrak{C}$ (usually seen as agents or coalitions), an attack relation (for modeling conflicts among elements of $\mathfrak{C}$), and a preference relation between elements of $\mathfrak{C}$ (to describe favorite agents/coalitions).

Definition 4 (Coalitional framework [3]). *A* coalitional framework *is a triple $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ where $\mathfrak{C}$ is a non-empty set of elements, $\mathcal{A} \subseteq \mathfrak{C} \times \mathfrak{C}$ is an attack relation, and $\prec$ is a preorder on $\mathfrak{C}$ representing preferences on elements in $\mathfrak{C}$.*

Let S be a non-empty set of elements. $\mathbb{CF}(S)$ denotes the set of all coalitional frameworks where elements are taken from the set S, i.e. for each $(\mathfrak{C}, \mathcal{A}, \prec) \in \mathbb{CF}(S)$ we have that $\mathfrak{C} \subseteq S$.

The set $\mathfrak{C}$ in Definition 4 is intentionally generic, accounting for various possible alternatives. One alternative is to consider $\mathfrak{C}$ as a set of agents $\mathbb{A}\text{gt} = \{a_1, \ldots, a_k\}$: $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec) \in \mathbb{CF}(\mathbb{A}\text{gt})$. Then, a *coalition* is given by $C = \{a_{i_1}, \ldots, a_{i_l}\} \subseteq \mathfrak{C}$ and "agent" can be used as an intuitive reference to elements of $\mathfrak{C}$. Another alternative is to use a coalitional framework $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ based on $\mathbb{CF}(\mathcal{P}(\mathbb{A}\text{gt}))$. Now elements of $\mathfrak{C} \subseteq \mathcal{P}(\mathbb{A}\text{gt})$ are *groups* or *coalitions* (where we consider singletons as groups too) of agents. Under this interpretation a coalition $C \subseteq \mathfrak{C}$ is a *set of sets* of agents. Although "coalition" is already used for $C \subseteq \mathfrak{C}$, we also use the intuitive reading "coalition" or "group" to address elements in $\mathfrak{C}$.[1] Yet another way is not to use the specific structure for elements in $\mathfrak{C}$, assuming it just consists of abstract elements, e.g. c_1, c_2, etc. One may think of these elements as individual agents or coalitions. This approach is followed in [3].

In the rest of this paper we mainly follow the first alternative when informally speaking about coalitional frameworks, i.e. we consider $\mathfrak{C}$ as a set of agents.

Example 1. *Consider the following two coalitional frameworks: (i) $\mathcal{CF}_1 = (\mathfrak{C}, \mathcal{A}, \prec)$ where $\mathfrak{C} = \{a_1, a_2, a_3\}$, $\mathcal{A} = \{(a_3, a_2), (a_2, a_1), (a_1, a_3)\}$ and agent a_3 is preferred over a_1, i.e. $a_1 \prec a_3$; and (ii) $\mathcal{CF}_2 = (\mathfrak{C}', \mathcal{A}', \prec')$ where $\mathfrak{C}' = \{\{a_1\}, \{a_2\}, \{a_3\}\}$, $\mathcal{A}' = \{(\{a_3\}, \{a_2\}), (\{a_2\}, \{a_1\}), (\{a_1\}, \{a_3\})\}$ and group $\{a_3\}$ is preferred over $\{a_1\}$, i.e. $\{a_1\} \prec'$*

[1]The first interpretation is a special case of the second (coalitional frameworks are members $\mathbb{CF}(\mathcal{P}(\mathbb{A}\text{gt}))$).

$\{a_3\}$. *They capture the same scenario and are isomorphic but* $\mathcal{CF}_1 \in \mathbb{CF}(\{a_1, a_2, a_3\})$ *and* $\mathcal{CF}_2 \in \mathbb{CF}(\mathcal{P}(\{a_1, a_2, a_3\}))$*; that is, the first framework is defined regarding single agents and the latter over (trivial) coalitions. Figure 1 (a) shows a graphical representation of the first coalitional framework.*

Let $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ be a coalitional framework. For $C, C' \in \mathfrak{C}$, we say that C *attacks* C' iff $C\mathcal{A}C'$. The attack relation represents conflicts between elements of $\mathfrak{C}$; for instance, two agents may rely on the same (unique) resource or they may have disagreeing goals, which prevents them from cooperation. However, the notion of attack may not be sufficient for modelling conflicts, as some elements (resp. coalitions) in $\mathfrak{C}$ may be preferred over others. This leads to the notion of *defeater* which combines the notions of attack and preference.

Definition 5 (Defeater). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework and let* $C, C' \in \mathfrak{C}$. *We say that* C defeats C' *if, and only if,* C *attacks* C' *and* C' *is not preferred over* C *(i.e., not* $C \prec C'$*). We also say that* C *is a* defeater *for* C'.

Attacks and defeats are defined between *single* elements of $\mathfrak{C}$. As we are interested in the formation of coalitions it is reasonable to consider conflicts between coalitions. Members in a coalition may prevent attacks to members in the same coalition; they protect each other. The concept of defence, introduced next, captures this idea of mutual protection.

Definition 6 (Defence). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework and* $C, C' \in \mathfrak{C}$. *We say that* C' defends itself against C *if, and only if,* C' *is preferred over* C*, i.e.,* $C \prec C'$*, and* C' defends itself *if it defends itself against any of its attackers. Furthermore,* C *is* defended by *a set* $\mathfrak{S} \subseteq \mathfrak{C}$ *of elements of* $\mathfrak{C}$ *if, and only if, for all* C' *defeating* C *there is a coalition* $C'' \in \mathfrak{S}$ *defeating* C'.

In other words, if an element C' defends itself against C then C may attack C' but C is not allowed to defeat C'.

A minimal requirement one should impose on a coalition is that its members do not defeat each other; otherwise, the coalition may be unstable and break up sooner or later because of conflicts among its members. This is formalized in the next definition.

Definition 7 (Conflict-free). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework and* $\mathfrak{S} \subseteq \mathfrak{C}$ *a set of elements in* $\mathfrak{C}$. *Then,* $\mathfrak{S}$ *is called* conflict-free *if, and only if, there is* no $C \in \mathfrak{S}$ *defeating some member of* $\mathfrak{S}$.

It must be remarked that our notions of "defence" and "conflict-free" are defined in terms of "defeat" rather than "attack".[2] Given a coalitional framework $\mathcal{CF}$ we will use argumentation to compute coalitions with desirable properties. In argumentation theory many different semantics have been proposed to define ultimately accepted arguments [10, 7]. We apply this rich framework to provide different ways to coalition formation. A semantics can be defined as follows.

Definition 8 (Coalitional framework semantics). *A* semantics *for a coalitional framework* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *is a (isomorphism invariant) mapping* $\mathfrak{sem}$ *which assigns to a given coalitional framework* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *a set of subsets of* $\mathfrak{C}$*, i.e.,* $\mathfrak{sem}(\mathcal{CF}) \subseteq \mathcal{P}(\mathfrak{C})$.

Let $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ be a coalitional framework. To formally characterize different semantics we will define a function $\mathcal{F}_{\mathcal{CF}} : \mathcal{P}(\mathfrak{C}) \to \mathcal{P}(\mathfrak{C})$ which assigns to a set of coalitions $\mathfrak{S} \in \mathcal{P}(\mathfrak{C})$ the coalitions defended by $\mathfrak{S}$.

Definition 9 (Characteristic function $\mathcal{F}$). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework and* $\mathfrak{S} \subseteq \mathfrak{C}$. *The function* $\mathcal{F}$ *defined by*

$$\mathcal{F}_{\mathcal{CF}} : \mathcal{P}(\mathfrak{C}) \to \mathcal{P}(\mathfrak{C})$$
$$\mathcal{F}_{\mathcal{CF}}(\mathfrak{S}) = \{C \in \mathfrak{C} \mid C \textit{ is defended by } \mathfrak{S}\}$$

is called characteristic function.[3]

$\mathcal{F}$ can be applied recursively to coalitions resulting in new coalitions. For example, $\mathcal{F}(\emptyset)$ provides all undefeated coalitions and $\mathcal{F}^2(\emptyset)$ constitutes the set of all elements of $\mathfrak{C}$ which members are undefeated *or* are defended by undefeated coalitions.

Example 2. *Consider again the coalitional framework* $\mathcal{CF}_1$ *given in Example 1. The characteristic function applied on the empty set results in* $\{a_3\}$ *since the agent is undefeated,* $\mathcal{F}(\emptyset) = \{a_3\}$. *Applying* $\mathcal{F}$ *on* $\mathcal{F}(\emptyset)$ *determines the set* $\{a_1, a_3\}$ *because* a_1 *is defended by* a_3. *It is easy to see that* $\{a_1, a_3\}$ *is a fixed point of* $\mathcal{F}$.

We now introduce the first concrete semantics called coalition structure semantics, which was originally defined in [3].

Definition 10 (Coalition structure $\mathfrak{sem}_{\textsc{cs}}$ [3]). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework. Then*

$$\mathfrak{sem}_{cs}(\mathcal{CF}) := \left\{ \bigcup_{i=1}^{\infty} \mathcal{F}^i_{\mathcal{CF}}(\emptyset) \right\}$$

is called coalition structure semantics *or just* coalition structure *for* $\mathcal{CF}$.

For a coalitional framework $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ with a finite set $\mathfrak{C}$[4] the characteristic function $\mathcal{F}$ is continuous [10, Lemma 28]. Since $\mathcal{F}$ is also monotonic it has a least fixed point given by $\mathcal{F}(\emptyset) \uparrow^{\omega}$ (according to Knaster-Tarski). We have the following straightforward properties of coalition structure semantics.

Proposition 1 (Coalition structure). *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework with a finite set* $\mathfrak{C}$. *There is always a unique coalition structure for* $\mathcal{CF}$. *Furthermore, if no element of* $C \in \mathfrak{C}$ *defends itself then the coalitional structure is empty, i.e.* $\mathfrak{sem}_{cs}(\mathcal{CF}) = \{\emptyset\}$.

Example 3. *The following situations illustrate the notion of coalitional structure:*

(a) Consider Example 2. Since $\{a_1, a_3\}$ *is a fixed point of* $\mathcal{F}_{\mathcal{CF}_1}$ *the coalitional framework* $\mathcal{CF}_1$ *has* $\{a_1, a_3\}$ *as coalitional structure.*

[2]In [3, 4] these notions are defined the other way around, resulting in a different characterization of stable semantics.

[3]We omit the subscript $\mathcal{CF}$ if it is clear from context.

[4]Actually, it is enough to assume that $\mathcal{CF}$ is finitary (cf. [10, Def. 27]).

(b) $\mathcal{CF}_3 := (\mathfrak{C}, \mathcal{A}, \prec) \in \mathbb{CF}(\{a_1, a_2, a_3\})$ *(shown in Figure 1(b)), is a coalitional framework with* $\mathfrak{C} = \{a_1, a_2, a_3\}$, $\mathcal{A} = \{(a_1, a_2), (a_1, a_3), (a_2, a_1), (a_2, a_3), (a_3, a_1)\}$ *and* a_3 *is preferred over* a_2, $a_2 \prec a_3$, *has the empty coalition as associated coalition str., i.e.* $\mathfrak{sem}_{cs}(\mathcal{CF}) = \{\emptyset\}$.

Since the coalition structure is often very restrictive, it seems reasonable to introduce other less restrictive semantics. Each of the following semantics are well-known in argumentation theory [10] and can be used as a criterium for coalition formation (cf. [3]).

Definition 11 (Argumentation Semantics). *Let* $(\mathfrak{C}, \mathcal{A}, \prec)$ *be a coalitional framework,* $\mathfrak{S} \subseteq \mathfrak{C}$ *a set of elements of* $\mathfrak{C}$. $\mathfrak{S}$ *is called*

(a) admissible extension *iff* $\mathfrak{S}$ *is conflict-free and* $\mathfrak{S}$ *defends all its elements, i.e.* $\mathfrak{S} \subseteq \mathcal{F}(\mathfrak{S})$.

(b) complete extension *iff* $\mathfrak{S}$ *is conflict-free and* $\mathfrak{S} = \mathcal{F}(\mathfrak{S})$.

(c) grounded extension *iff* $\mathfrak{S}$ *is the smallest (wrt. to set inclusion) complete extension.*

(d) preferred extension *iff* $\mathfrak{S}$ *is a maximal (wrt. to set inclusion) admissible extension.*

(e) stable extension *iff* $\mathfrak{S}$ *is conflict-free and it defeats all arguments not in* $\mathfrak{S}$.

Let $\mathfrak{sem}_{cs}$ *(resp.* $\mathfrak{sem}_{complete}$, $\mathfrak{sem}_{grounded}$, $\mathfrak{sem}_{preferred}$ *and* $\mathfrak{sem}_{stable}$*) denote the semantics which assigns to a coalitional structure* $\mathcal{CF}$ *all its admissible (resp. complete, grounded, preferred, and stable) extensions.*

There is only one unique coalition structure (possibly the empty one) for a given coalitional framework, but there can be several stable and preferred extensions. The existence of at least one preferred extension is assured which is not the case for the stable semantics. Thus, the possible coalitions very much depend on the used semantics.

Example 4. *For* $\mathcal{CF}_3$ *from Example 3 the following holds:*

$$\mathfrak{sem}_{cs}(\mathcal{CF}) = \{\emptyset\}$$
$$\mathfrak{sem}_{admissible}(\mathcal{CF}) = \{\{a_1\}, \{a_2\}, \{a_3\}, \{a_2, a_3\}\}$$
$$\mathfrak{sem}_{complete}(\mathcal{CF}) = \mathfrak{sem}_{grounded}(\mathcal{CF}) = \{\{a_1\}, \{a_2, a_3\}\} =$$
$$\mathfrak{sem}_{preferred}(\mathcal{CF}) = \mathfrak{sem}_{stable}(\mathcal{CF}) = \{\{a_1\}, \{a_2, a_3\}\}$$

Analogously, for the coalitional framework $\mathcal{CF}_1$ *from Example 1 there exists one complete extension* $\{a_1, a_3\}$ *which is also a grounded, preferred, and stable extension.*

4. COALITIONAL ATL

In this section we combine *argumentation for coalition formation* and **ATL** and introduce *coalitional* ***ATL*** ($\mathbf{ATL^c}$). This logic extends **ATL** by new operators $\langle\!| A |\!\rangle$ for each subset $A \subseteq \mathbb{A}\mathrm{gt}$ of agents. These new modalities combine, or rather integrate, coalition formation into the original **ATL** cooperation modalities $\langle\!\langle A \rangle\!\rangle$. The intended reading of $\langle\!| A |\!\rangle\varphi$ is that the group A of agents *is able to form a coalition* $B \subseteq \mathbb{A}\mathrm{gt}$ *such that* A *is a part of* B, $A \subseteq B$, and B *can enforce* φ. Coalition formation is modeled by the formal argumentative approach in the context of coalition formation, as described in Section 3, based on the framework developed in [3].

Our main motivation for this logic is to make it possible to reason about the ability of building coalition structures, and not only about an *a priori* specified group of agents (as it is the case for $\langle\!\langle A \rangle\!\rangle\varphi$). The new modality $\langle\!| A |\!\rangle$ provides a rather subjective view of the agents in A and their power to create a supergroup B, $A \subseteq B$, which in turn is used to reason about the ability to enforce a given property.

The language of $\mathbf{ATL^c}$ is as follows.

Definition 12 ($\mathcal{L}_{ATL^c}$). *Let* $\mathbb{A}\mathrm{gt} = \{a_1, \ldots, a_k\}$ *be a finite, nonempty set of agents, and* Π *be a set of propositions (with typical element* p*). We use the symbol "*a*" to denote a typical agent, and "*A*" to denote a typical group of agents from* $\mathbb{A}\mathrm{gt}$*. The logic* $\mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$ *is defined by the following grammar:*

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\!\langle A \rangle\!\rangle \bigcirc \varphi \mid \langle\!\langle A \rangle\!\rangle \Box\, \varphi \mid \langle\!\langle A \rangle\!\rangle \varphi \mathcal{U} \varphi \mid$$
$$\langle\!| A |\!\rangle \bigcirc \varphi \mid \langle\!| A |\!\rangle \Box\, \varphi \mid \langle\!| A |\!\rangle \varphi \mathcal{U} \varphi$$

We extend concurrent game structures by *coalitional frameworks* and an *argumentative semantics.* A coalitional framework is assigned to each state of the model capturing the current "conflicts" among agents. In doing so, we allow that conflicts change over time, being thus *state dependent.* But we assume that the argumentative semantics is the same for all states.

Definition 13 (**CCGS**). *A* coalitional concurrent game structure *(**CCGS**) is given by a tuple*

$$\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem} \rangle$$

where $\langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o \rangle$ *is a* ***CGS****,* $\zeta : Q \to \mathbb{CF}(\mathbb{A}\mathrm{gt})$ *is a function which assigns a coalitional framework over* $\mathbb{A}\mathrm{gt}$ *to each state of the model, and* $\mathfrak{sem}$ *is an (argumentative) semantics as defined in Definition 8.*

A model provides an argumentation semantics $\mathfrak{sem}$ which assigns all formable coalitions to a given coalitional framework. As argued before we require from a valid coalition that it is not only justified by the argumentation semantics but that it is also a superset of the predetermined starting coalition. This leads to the notion *valid coalition.*

Definition 14 (Valid coalition). *Let* A, $B \subseteq \mathbb{A}\mathrm{gt}$ *be groups of agents,* $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem} \rangle$ *be a* ***CCGS*** *and* $q \in Q$.

We say that B *is a* valid coalition *with respect to* A*,* q*, and* $\mathcal{M}$ *whenever* $B \in \mathfrak{sem}(\zeta(q)) \cup \{A\}$ *and* $A \subseteq B$*. Furthermore, we use* $\boldsymbol{VC}_{\mathcal{M}}(A, q)$ *to denote the set of all valid coalitions regarding* A*,* q*, and* $\mathcal{M}$*. The subscript* $\mathcal{M}$ *is omitted if clear from the context.*

Consider a formula $\langle\!| A |\!\rangle\psi$. Since we assume that the members of the initial group A work together, whatever the reasons might be, group A is added to $\mathfrak{sem}(\zeta(q))$. This ensures that agents in A can enforce ψ on their own, if they are able to do so, even if A is not accepted originally by the argumentation semantics, i.e. $A \notin \mathfrak{sem}(\zeta(q))$.

There are other sensible ways of defining a valid coalition. For instance, one may require that $B \subseteq A$ instead of $A \subseteq B$. In this scenario members from the "initial coalition" A are removed to find smaller coalitions that ensure a property. It can be thought of reducing costs of the overall coalition and to increase the payoff of the coalition's members. Another arguable point is whether the indicated initial coalition A should always be considered as a valid coalition. However, in the rest of the paper, due to the lack of space, we will stick to our initial motivation and leave alternatives for future research.

The semantics of the new modality is given by

Definition 15 (**ATL^c** Semantics). *Let a **CCGS** $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem}\rangle$, $A \subseteq \mathbb{A}\mathrm{gt}$ a group of agents and $q \in Q$ be given. The semantics of coalitional **ATL** extends that of **ATL**, given in Definition 3, by the following rule ($\langle\!| A |\!\rangle\psi \in \mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$):*

$\mathcal{M}, q \models \langle\!| A |\!\rangle\psi$ *iff there is a coalition* $B \in \boldsymbol{VC}(A, q)$ *such that* $\mathcal{M}, q \models \langle\!\langle B \rangle\!\rangle\psi$.

Remark 2 (Different Semantics, $\models_{\mathfrak{sem}}$). *We have just defined a whole class of semantic rules for modality $\langle\!| \cdot |\!\rangle$. The actual instantiation of the semantics $\mathfrak{sem}$, for example $\mathfrak{sem}_{stable}$, $\mathfrak{sem}_{pref}$, and $\mathfrak{sem}_{cs}$ defined in Section 3, affects the semantics of the cooperation modality.*

*For the sake of readability, we sometimes annotate the satisfaction relation $\models$ with the presently used argumentation semantics. That is, given a **CCGS** $\mathcal{M}$ with an argumentation semantics $\mathfrak{sem}$ we write $\models_{\mathfrak{sem}}$ instead of $\models$.*

The underlying idea of the semantic definition of $\langle\!| A |\!\rangle\psi$ is as follows. A given (initial) group of agents $A \subseteq \mathbb{A}\mathrm{gt}$ is able to form a *valid coalition* B (i.e. A is part of the coalition B and $A \subseteq B$), with respect to a given coalitional framework $\mathcal{CF}$ and a particular semantics $\mathfrak{sem}$, such that B can enforce ψ.

Similarly to the alternatives to our definition of valid coalitions there are other sensible semantics for **ATL^c**. The semantics we presented here is not particularly dependent on time; i.e., except from the selection of a valid coalition B at the initial state there is no further interaction between time and coalition formation. We have chosen this simplistic definition to present our main idea – the connection of **ATL** and coalition formation by means of argumentation – as clear as possible. We will briefly discuss alternative semantics in Section 6, that are worth to be investigated in the future.

A simple relation between the classic and the new modality is given next.

Proposition 3. *Let $A \subseteq \mathbb{A}\mathrm{gt}$ and $\langle\!\langle A \rangle\!\rangle\psi \in \mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$. Then it holds that the formula $\langle\!\langle A \rangle\!\rangle\psi \rightarrow \langle\!| A |\!\rangle\psi$ is valid[5] with respect to **CCGS**'s.*

Note that $\langle\!| A |\!\rangle\varphi$ does not necessarily imply $\langle\!| A' |\!\rangle\varphi$ when $A' \neq A$. In the first formula A is taken as a valid coalition (by definition) which is not always the case in the latter formula.

Proposition 4. *Let $A \subseteq \mathbb{A}\mathrm{gt}$ and $\langle\!| A |\!\rangle\psi \in \mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$. Then it holds that $\langle\!| A |\!\rangle\psi \rightarrow \bigvee_{B \in \mathcal{P}(\mathbb{A}\mathrm{gt}), A \subseteq B} \langle\!\langle B \rangle\!\rangle\psi$ is a validity with respect to **CCGS**'s.*

Compared to **ATL**, a formula like $\langle\!| A |\!\rangle\varphi$ does *not* refer to the ability of A to enforce φ, but rather to the ability of A to *constitute* a coalition B, such that $A \subseteq B$, and then, in a second step, to the ability of B to enforce φ.

[5]That is, the formula holds in all states of the given model for any **CCGS**.

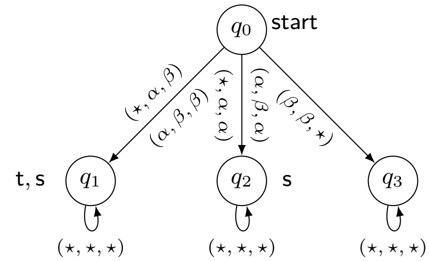

Figure 2: A simple CGS defined in Example 5.

Thus, two different notions of ability are captured in these new modalities. For instance, $\langle\!\langle A \rangle\!\rangle\psi \wedge \neg\langle\!| \emptyset |\!\rangle\psi$ expresses that group A of agents can enforce ψ, but there is no *reasonable* coalition which can enforce ψ (particularly not A, although they possess the theoretical power to do so).

Example 5. *There are three agents a_1, a_2, and a_3 which prefer different outcomes. Agent a_1 (resp. a_2, a_3) desires to get outcome r (resp. s, t). One may assume that all outcomes are distinct; for instance, a_1 is not satisfied with an outcome x whenever $x \neq \mathsf{r}$. Each agent can choose to perform action α or β. Action profiles and their outcomes are shown in Figure 2. The $\star$ is used as a placeholder for any of the two actions, i.e. $\star \in \{\alpha, \beta\}$. For instance, the profile $(\beta, \beta, \star)$ leads to state q_3 whenever agent a_1 and a_2 perform action β and a_3 either does α or β.*

According to the scenario depicted in the figure, a_1 and a_2 cannot commonly achieve their goals. The same holds for a_1 and a_3. On the other hand, there exists a situation, q_1, in which both agents a_2 and a_3 are satisfied. One can formalize the situation as the coalitional game $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec)$ given in Example 3(b), that is, $\mathfrak{C} = \mathbb{A}\mathrm{gt}$, $\mathcal{A} = \{(a_1, a_2), (a_1, a_3), (a_2, a_1), (a_2, a_3), (a_3, a_1)\}$ and $a_2 \prec a_3$.

*We formalize the example as the **CCGS** $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem}\rangle$ where $\mathbb{A}\mathrm{gt} = \{a_1, a_2, a_3\}$, $Q = \{q_0, q_1, q_2, q_3\}$, $\Pi = \{\mathsf{r}, \mathsf{s}, \mathsf{t}\}$, and $\zeta(q) = \mathcal{CF}$ for all states $q \in Q$. Transitions and the state labeling can be seen in Figure 2. Furthermore, we do not specify a concrete semantics $\mathfrak{sem}$ yet, and rather adjust it in the remainder of the example.*

*We can use pure **ATL** formulas, i.e. formulas not containing the new modalities $\langle\!| \cdot |\!\rangle$, to express what groups of agents can achieve. We have, for instance, that agents a_1 and a_2 can enforce a situation which is undesirable for a_3: $\mathcal{M}, q_0 \models \langle\!\langle a_1, a_2 \rangle\!\rangle \bigcirc \mathsf{r}$. Indeed, $\{a_1, a_2\}$ and the grand coalition $\mathbb{A}\mathrm{gt}$ (since it contains $\{a_1, a_2\}$) are the only coalitions which are able to enforce $\bigcirc\mathsf{r}$; we have*

$$\mathcal{M}, q_0 \models \neg\langle\!\langle X \rangle\!\rangle \bigcirc \mathsf{r} \qquad (1)$$

for all $X \subset \mathbb{A}\mathrm{gt}$ and $X \neq \{a_1, a_2\}$. Outcomes s or t can be enforced by a_2: $\mathcal{M}, q_0 \models \langle\!\langle a_2 \rangle\!\rangle \bigcirc(\mathsf{s} \vee \mathsf{t})$. Agents a_2 and a_3 also have the ability to enforce a situation which agrees with both of their desired outcomes: $\mathcal{M}, q_0 \models \langle\!\langle a_2, a_3 \rangle\!\rangle \bigcirc(\mathsf{s} \wedge \mathsf{t})$ Furthermore, agent a_2 can decide on its preferred outcome s on its own: $\mathcal{M}, q_0 \models \langle\!\langle a_2 \rangle\!\rangle \bigcirc\mathsf{s}$

These properties do not take into account the coalitional framework, that is whether specific coalitions can be formed or not. By using the coalitional framework, we get

$$\mathcal{M}, q_0 \models_{\mathfrak{sem}} \langle\!\langle a_1, a_2 \rangle\!\rangle \bigcirc \mathsf{r} \wedge \neg\langle\!| a_1 |\!\rangle \bigcirc \mathsf{r} \wedge \neg\langle\!| a_2 |\!\rangle \bigcirc \mathsf{r}$$

for any semantics $\mathfrak{sem}$ introduced in Definition 8 and calculated in Example 4. The possible coalition (resp. coalitions) containing a_1 (resp. a_2) is $\{a_1\}$ (resp. are $\{a_2\}$ and

$\{a_2, a_3\}$). *But neither of these can enforce* $\bigcirc \mathsf{r}$ *(in* q_0*) because of (1). Thus, although it is the case that the coalition* $\{a_1, a_2\}$ *has the theoretical ability to enforce* r *in the next moment (which is a "losing" situation for* a_3*),* a_3 *should not consider it as sensible since agents* a_1 *and* a_2 *would not agree to constitute a coalition (according to the coalitional framework* $\mathcal{CF}$*).*

The decision for a specific semantics is a crucial point and depends on the actual application. The next example shows that with respect to a particular argumentation semantics, agents are able to form a coalition which can successfully achieve a given property, whereas another argumentative semantics does not allow that.

EXAMPLE 6. ***ATL***$^{\mathbf{c}}$ *can be used to determine whether a coalition for enforcing a specific property exists. Assume that* $\mathfrak{sem}$ *represents the grounded semantics. For instance, the statement*

$$\mathcal{M}, q_0 \models_{\mathfrak{sem}_{grounded}} \langle\!|\emptyset|\!\rangle \bigcirc \mathsf{t}$$

expresses that there is a grounded extension (i.e. a coalition wrt to the grounded semantics) which can enforce $\bigcirc\mathsf{t}$*, namely the coalition* $\{a_2, a_3\}$*. This result does not hold for all semantics; for instance, we have*

$$\mathcal{M}, q_0 \not\models_{\mathfrak{sem}_{cs}} \langle\!|\emptyset|\!\rangle \bigcirc \mathsf{t}$$

with respect to the coalition structure semantics, since the coalition structure is the empty coalition and $\mathcal{M}, q_0 \not\models \langle\!\langle\emptyset\rangle\!\rangle \bigcirc \mathsf{t}$.

Note that it is easily possible to extend the language by an *update mechanism*, in order to compare different argumentative semantics using formulae inside the object language.

5. MODEL CHECKING ATL$^{\mathbf{c}}$

In this section we present an algorithm for model checking **ATL**$^{\mathbf{c}}$ formulae. The model checking problem is given by the question whether a given **ATL**$^{\mathbf{c}}$ formula follows from a given **CCGS** $\mathcal{M}$ and state q, i.e. whether $\mathcal{M}, q \models \varphi$ [9]. In [2] it is shown that model checking **ATL** is **P**-complete, with respect to the number of transitions of $\mathcal{M}$, m, and the length of the formula, l, and can be done in time $\mathcal{O}(m \cdot l)$.

For **ATL**$^{\mathbf{c}}$ we also have to treat the new coalitional modalities in addition to the normal **ATL** constructs. Let us consider the formula $\langle\!|A|\!\rangle\psi$. According to the semantics of $\langle\!|A|\!\rangle$, given in Definition 15, we must check whether there is a coalition B such that (i) $A \subseteq B$, (ii) B is acceptable by the argumentation semantics or $A = B$, and (iii) $\langle\!\langle B\rangle\!\rangle\psi$. The number of possible candidate coalitions B which satisfy (i) and (ii) is bounded by $|\mathcal{P}(\mathbb{A}\mathrm{gt})|$. Thus, in the worst case there might be *exponentially* many calls to a procedure checking whether $\langle\!\langle B\rangle\!\rangle\psi$. Another source of complexity is the time needed to compute the argumentation semantics. In [11], for instance, it is stated that credulous acceptance[6] using preferred semantics is **NP**-complete.

Both considerations together suggest that the model checking complexity has two computationally hard parts: exponentially many calls to $\langle\!\langle A\rangle\!\rangle\psi$ and the computation of the argumentation semantics. Indeed, Theorem 6 will support this intuition. However, we show that it is possible to "combine" both computationally hard parts to obtain an algorithm which is in $\mathbf{\Delta_2^P} = \mathbf{P^{NP}}$, if the computational complexity to determine whether a given coalition is acceptable are not harder than **NP**.

For the rest of this Section, we will denote by $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}}$ the set of all coalitions A such that A is acceptable according to the coalitional framework $\mathcal{CF}$ and the argumentation semantics $\mathfrak{sem}$, i.e. $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} := \{A \mid A \in \mathfrak{sem}(\mathcal{CF})\}$.

Given some complexity class $\mathcal{C}$, we use the notation $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in$

function *mcheck*$(\mathcal{M}, q, \varphi)$;
Given a **CCGS** $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem}\rangle$, a state $q \in Q$, and $\varphi \in \mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$ the algorithm returns $\top$ if, and only if, $\mathcal{M}, q \models_{\mathfrak{sem}} \varphi$.

case φ **contains no** $\langle\!|B|\!\rangle$**: if** $q \in mcheck_{\mathbf{ATL}}(\mathcal{M}, \varphi)$ **return** $\top$ **else** $\bot$

case φ **contains some** $\langle\!|B|\!\rangle$**:**

- **case** $\varphi \equiv \neg\psi$**: return** $\neg(\mathcal{M}, q, \psi)$
- **case** $\varphi \equiv \psi \vee \psi'$**: return** $mcheck(\mathcal{M}, q, \psi) \vee mcheck(\mathcal{M}, q, \psi')$
- **case** $\varphi \equiv \langle\!\langle A\rangle\!\rangle T\psi$**:** Label all states q' where $mcheck(\mathcal{M}, q', \psi) == \top$ with a new proposition yes and return $mcheck(\mathcal{M}, q, \langle\!\langle A\rangle\!\rangle T\mathsf{yes})$; T stands for $\Box$ or $\bigcirc$.
- **case** $\varphi \equiv \langle\!\langle A\rangle\!\rangle \psi\, \mathcal{U}\, \psi'$**:** Label all states q' where $mcheck(\mathcal{M}, q', \psi) == \top$ with a new proposition yes_1, all states q' where $mcheck(\mathcal{M}, q', \psi') == \top$ with a new proposition yes_2 and return $mcheck(\mathcal{M}, q, \langle\!|A|\!\rangle \mathsf{yes}_1\, \mathcal{U}\, \mathsf{yes}_2)$
- **case** $\varphi \equiv \langle\!|A|\!\rangle T\psi$**,** ψ **contains some** $\langle\!|C|\!\rangle$**:** Label all states q' where $mcheck(\mathcal{M}, q', \psi) == \top$ with a new proposition yes and return $mcheck(\mathcal{M}, q, \langle\!|A|\!\rangle T\mathsf{yes})$; T stands for $\Box$ or $\bigcirc$.
- **case** $\varphi \equiv \langle\!|A|\!\rangle \psi\, \mathcal{U}\, \psi'$**,** ψ **or** ψ' **contain some** $\langle\!|C|\!\rangle$**:** Label all states q' where $mcheck(\mathcal{M}, q', \psi) == \top$ with a new proposition yes_1, all states q' where $mcheck(\mathcal{M}, q', \psi') == \top$ with a new proposition yes_2 and return $mcheck(\mathcal{M}, q, \langle\!|A|\!\rangle \mathsf{yes}_1\, \mathcal{U}\, \mathsf{yes}_2)$
- **case** $\varphi \equiv \langle\!|A|\!\rangle\psi$ **and** ψ **contains no** $\langle\!|C|\!\rangle$**:** Non-deterministically choose $B \in \mathcal{P}(\mathbb{A}\mathrm{gt})$
  if
  (1) $B \in (\mathfrak{sem}(\zeta(q)) \cup \{A\})$,
  (2) $A \subseteq B$, and $\quad (\star)$
  (3) $q \in mcheck_{\mathbf{ATL}}(\mathcal{M}, \langle\!\langle B\rangle\!\rangle\psi)$
  then return $\top$ **else** $\bot$

function $mcheck_{\mathbf{ATL}}(\mathcal{M}, \varphi)$;
Given a **CGS** $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o\rangle$ and $\varphi \in \mathcal{L}_{ATL}(\mathbb{A}\mathrm{gt}, \Pi)$, the standard **ATL** model checking algorithm (cf. [2]) returns all states q with $\mathcal{M}, q \models_{\mathbf{ATL}} \varphi$.

- return $\{q \in Q \mid \mathcal{M}, q \models_{\mathbf{ATL}} \varphi\}$

Figure 3: A model checking algorithm for ATL$^{\mathbf{c}}$

[6]That is, whether an argument is in *some* preferred extension.

$\mathcal{C}$ to state that the word problem of $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}}$, i.e., whether a A is a member of $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}}$, is in $\mathcal{C}$.

Before we turn to the model checking algorithm, we introduce the following result.

LEMMA 5. *Let* $\mathcal{CF} = (\mathfrak{C}, \mathcal{A}, \prec) \in \mathbb{CF}(\mathbb{A}\mathrm{gt})$ *be a **CCGS**. For all semantics* $\mathfrak{sem}$ *defined in Definition 11 we have that* $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in \mathbf{P}$.

In Figure 3 we propose a model checking algorithm for $\mathbf{ATL^c}$. The complexity result given in the next theorem is modulo the complexity needed to compute membership in $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}}$.

THEOREM 6 (MODEL CHECKING $\mathbf{ATL^c}$). *Let a **CCGS*** $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem}\rangle$ *be given,* $q \in Q$, $\varphi \in \mathcal{L}_{ATL^c}(\mathbb{A}\mathrm{gt}, \Pi)$, *and* $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in \mathcal{C}$. *Model checking* $\boldsymbol{ATL^c}$ *with respect to the argumentation semantics* $\mathfrak{sem}$[7] *is in* $\mathbf{P}^{\mathbf{NP}^{\mathcal{C}}}$.

The last theorem gives an upper bound for model checking $\mathbf{ATL^c}$ with respect to an arbitrary but fixed semantics $\mathfrak{sem}$. A finer grained classification of the computational complexity of $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}}$ allows to improve the upper bound given in Theorem 6. Assume that $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in \mathbf{P}$ and consider the last case of function *mcheck* in Figure 3 labelled by $(\star)$, $\varphi \equiv \langle\!|A|\!\rangle\psi$. First, a coalition $B \in \mathcal{P}(\mathbb{A}\mathrm{gt})$ is non-deterministically chosen and then, it is checked whether B satisfies the three conditions (1-3) in $(\star)$. Each of the three tests can be done in deterministic polynomial time. Hence, the verification of $\mathcal{M}, q \models \langle\!|A|\!\rangle\psi$, in the last case, meets the "guess and verify" principle which is characteristic for problems in $\mathbf{NP}$. This brings the overall complexity of the algorithm to $\mathbf{\Delta_2^P}$. More surprisingly, the same result holds even for the case where $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in \mathbf{NP}$.

COROLLARY 7. *If* $\mathcal{L}_{\mathfrak{sem},\mathcal{CF}} \in \mathbf{NP}$ *(resp.* $\mathbf{NP}$*-complete) then model checking* $\boldsymbol{ATL^c}$ *is in* $\mathbf{\Delta_2^P}$ *(resp.* $\mathbf{\Delta_2^P}$*-complete) with respect to* $\mathfrak{sem}$.

In [11] the complexity of credulous reasoning with respect to the preferred and stable extensions is analyzed and determined to be $\mathbf{NP}$-complete. This is in the line with our result: there can be a polynomial number of calls to $mcheck(\mathcal{M}, q, \langle\!|A|\!\rangle\psi)$ (where ψ does not contain any cooperation modality $\langle\!| \cdot |\!\rangle$). Now, the problem of checking whether $mcheck(\mathcal{M}, q, \langle\!|A|\!\rangle\psi)$ holds is very similar to checking whether some argument is credulously accepted. In both cases we have to ask for the existence of a set X with specific properties (in our framework we refer to X as a coalition and in [11] as an argument) which can be validated in polynomial deterministic time.

COROLLARY 8. *Model checking* $\boldsymbol{ATL^c}$ *is in* $\mathbf{\Delta_2^P}$ *for all semantics defined in Definition 11.*

6. ALTERNATIVE SEMANTICS

In this section we discuss two other semantics for the coalitional operators. The general idea is to intensify the interplay between the coalition formation process and the temporal structure of the model.

In the semantics presented in Definition 15 a valid coalition is initially formed and kept until the property is fulfilled.

[7]That is, whether $\mathcal{M}, q \models_{\mathfrak{sem}} \varphi$.

For instance, consider formula $\langle\!|A|\!\rangle\Box\,\varphi$. The formula is true in q if a valid coalition B *in* q can be formed such that it can ensure $\Box\,\varphi$. On might strengthen the scenario and require that B must be valid in *each* state on the path λ satisfying φ. Formally, the semantics could be given as follows: $q \models \langle\!|A|\!\rangle\Box\,\varphi$ if, and only if, $q \models \varphi$ and there is a coalition $B \in \mathbf{VC}(q, A)$ and a common strategy $s_B \in \Sigma_B$ such that for all paths $\lambda \in out(q, s_B)$ and for all $i \in \mathbb{N}$ it holds that $\lambda[i] \models \varphi$ *and* $B \in \mathbf{VC}(\lambda[i], A)$. The last part specifies that B must be a valid coalition in each state $q_i = \lambda[i]$ of λ.

In the semantics just presented the formed coalition B must persist over time until φ is enforced. One can go one step further. Instead of keeping the same coalition B it can also be sensible to consider "new" valid coalitions in each time step (wrt. A), possibly distinct from B. This leads to some kind of fixed point definition: $q \models \langle\!|A|\!\rangle\Box\,\varphi$ if, and only if, there is a coalition $B \in \mathbf{VC}(q, A)$ such that $q \models \varphi$ and $q \models \langle\!\langle B\rangle\!\rangle \bigcirc \langle\!|A|\!\rangle\Box\,\varphi$. At first, B must be a valid coalition in state q leading to a state in which φ is fulfilled and in which another valid coalition (wrt. A and the new state) exists which in turn can ensure to enter a state in which, again, there is another valid coalition and so on. The semantics of the remaining temporal operators are given analogously.

Another issue we would like to mention is the incorporation of our framework into $\mathbf{ATL}^*$. The semantics chosen in this paper (see Definition 15) can directly be transferred to $\mathbf{ATL}^*$. The adoption of the other discussed semantics is not that straightforward. It seems necessary to keep track of the coalition of interest. In the case of the first alternative semantics discussed in this section this would mean to recall B and in the second A. This could be done by annotating the satisfaction relation by a coalition.

In our future research we would like to consider more sophisticated semantics, investigate their effects on the computational complexity results, and analyze the interplay between different semantics.

7. RELATED WORK

To the best of our knowledge, there is no similar work of integrating a temporal logic like **ATL** and argument-based coalition formation. The main inspiration for our work was the powerful argumentation-based model for reasoning about coalition structures proposed by Amgoud [3]. Indeed, our notion of coalitional framework (Def. 4) is based on the notion of framework for generating coalition structures (FCS) presented in Amgoud's paper. However, our work is concerned with extending **ATL** by argumentation in order to model coalition formation.

In [12] an argumentation-based negotiation method for coalition formation is proposed, combining a logical framework and an argument evaluation mechanism. The proposed system involves several user agents and a mediator agent. During the negotiation, the mediator agent encourages appropriate user agents to join in a coalition in order to facilitate reaching an agreement. User agents advance proposals using a part of the user´s valuations in order to reflect the user´s preferences in an agreement. This approach differs greatly from our proposal, as we are not concerned with the negotiation process among agents, and our focus is on modelling coalitions within an extension of a highly expressive temporal logic, where coalition formation is part of the logical language.

Recent research in formalizing coalition formation has been oriented towards adding more expressivity to Pauly's coalition logic [17]. E.g. in [1], the authors define *Quantified Coalition Logic*, extending coalition logic with a limited but useful form of quantification to express properties such as *"there exists a coalition C satisfying property P such that C can achieve ϕ"*. In [5], a semantic translation from coalition logic to a fragment of an action logic is defined, connecting the notions of coalition power and the actions of the agents. However, in none of these cases argumentation is used to model the notion of coalition formation as done in this paper.

8. CONCLUSIONS

In this paper we have presented $\mathcal{L}_{ATL^c}$, a formal extension of **ATL** which is able to model coalition formation through argumentation. $\mathbf{ATL^c}$ does contain two different modalities, $\langle\langle A\rangle\rangle$ and $\langle\!| A |\!\rangle$, which refer to abilities of agents. $\langle\langle A\rangle\rangle$ is used to reason about the pure ability of the very group A. The question whether it is reasonable to assume that the members of A collaborate is not taken into account in **ATL**. With the new operator we try to close this gap and also allow to focus on *sensible coalition structures*. Here, "sensible" refers to *acceptable coalitions with respect to some argumentative semantics* (characterized in Definition 8).

We have defined the formal machinery required for characterizing argument-based coalition formation in terms of a new construct. The actual computation of the coalition is modeled in terms of a given argumentation semantics which can easily be changed in the model. This allows us to compare the ability of agents to form particular coalitions and study emerging properties regarding different semantics. As outlined in Section 5, the model checking used in **ATL** can be extended to $\mathcal{L}_{ATL^c}$ by integrating suitable proof procedures for argumentation semantics.

Recently, the adequate formalization of preferences has deserved considerable attention within the argumentation community, particularly in the context of the work of Kaci et al. [15]. Indeed, one of our future research lines is to extend our current formalization of $\mathbf{ATL^c}$ to capture more complex issues in preference handling and to consider more sophisticated semantics as discussed in Section 6.

Part of our future work also involves the actual implementation of a subset of $\mathcal{L}_{ATL^c}$, restricted to some particular argumentation semantics for which proof procedures can be easily deployed (e.g. grounded semantics), in order to perform experiments to assess our proposal when modeling complex problems. Research in this direction is currently being pursued.

Acknowledgement. This research was funded by the projects DAAD-SeCyT (DA0609) and PGI-UNS (24/ZN10). The authors would also like to thank the anonymous reviewers for their useful comments.

9. REFERENCES

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. Quantified coalition logic. In *IJCAI*, pages 1181–1186, 2007.

[2] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. *Journal of the ACM*, 49:672–713, 2002.

[3] L. Amgoud. An argumentation-based model for reasoning about coalition structures. In *ArgMAS*, pages 217–228, 2005.

[4] L. Amgoud. Towards a formal model for task allocation via coalition formation. In *AAMAS*, pages 1185–1186, 2005.

[5] S. Borgo. Coalitions in action logic. In *IJCAI*, pages 1822–1827, 2007.

[6] N. Bulling and W. Jamroga. A logic for reasoning about rational agents: Yet another attempt. In L. Czaja, editor, *Proceedings of CS&P*, 2007.

[7] M. Caminada. Semi-stable semantics. In *Intl. Conference on Computational Models of Argument (COMMA)*, pages 121–130, 2006.

[8] C. I. Chesñevar, A. G. Maguitman, and R. P. Loui. Logical models of argument. *ACM Comput. Surv.*, 32(4):337–383, 2000.

[9] E. Clarke, O. Grumberg, and D. Peled. *Model Checking*. MIT Press, 1999.

[10] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[11] P. E. Dunne and T. J. M. Bench-Capon. Two party immediate response disputes: Properties and efficiency. *Artif. Intell.*, 149(2):221–250, 2003.

[12] H. Hattori, T. Ito, T. Ozono, and T. Shintani. An approach to coalition formation using argumentation-based negotiation in multi-agent systems. In *Proceedings of IEA/AIE*, pages 687–696, London, UK, 2001. Springer-Verlag.

[13] W. Jamroga and N. Bulling. A general framework for reasoning about rational agents. In *Proceedings of AAMAS'07*, pages 592–594, Honolulu, Hawaii, USA, 2007. ACM Press. Short paper.

[14] W. Jamroga and N. Bulling. A logic for reasoning about rational agents. In F. Sadri and K. Satoh, editors, *Proceedings of CLIMA '07*, pages 54–69, Porto, Portugal, 2007. Univesidade Do Porto.

[15] S. Kaci, L. W. N. van der Torre, and E. Weydert. On the acceptability of incompatible arguments. In *Proc. of ECSQARU Conf.*, pages 247–258, 2007.

[16] N. C. Karunatillake, N. R. Jennings, I. Rahwan, and S. D. Ramchurn. Managing social influences through argumentation-based negotiation. In *Proc. of AAMAS*, pages 426–428, 2006.

[17] M. Pauly. A modal logic for coalitional power in games. *J. Log. Comput.*, 12(1):149–166, 2002.

[18] H. Prakken and G. Vreeswijk. Logical Systems for Defeasible Argumentation. In D. Gabbay and F.Guenther, editors, *Handbook of Phil. Logic*, pages 219–318. Kluwer, 2002.

[19] I. Rahwan and L. Amgoud. An argumentation based approach for practical reasoning. In *Proc. of AAMAS Conf.*, pages 347–354, 2006.

[20] I. Rahwan, P. Pasquier, L. Sonenberg, and F. Dignum. On the benefits of exploiting underlying goals in argument-based negotiation. In *Proc. of AAAI Conf.*, pages 116–121, 2007.

[21] Y. Tang and S. Parsons. Argumentation-based dialogues for deliberation. In *Proc. of AAMAS*, pages 552–559, 2005.