Modelling and Verifying Coalitions using Argumentation and ATL

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

Abstract During the last decade argumentation has evolved as a successful approach to formalize commonsense reasoning and decision making in multiagent systems. In particular, recent research has shown that argumentation can be used to provide a suitable framework for reasoning about coalition formation: which coalitions can be formed using different argumentation semantics. At the same time Alternating-time Temporal Logic (ATL for short) has been successfully used to reason about the behavior and abilities of coalitions of agents. However, an important limitation of ATL operators is that they account only for the existence of successful strategies of coalitions, not considering whether coalitions can be actually formed.

This paper is an attempt to combine both frameworks in order to develop a logical system through which we can reason at the same time (1) about abilities of coalitions of agents and (2) about the formation of coalitions. In order to achieve this, we provide a formal extension of ATL, called Coalitional ATL (CoalATL for short), in which the actual computation of the coalition is modelled in terms of argumentation semantics. Moreover, we integrate goals as agents’ incentive to join coalitions and examine the model checking complexity. Particularly, we show that model checking CoalATL is \( \Delta^2_P \)-complete in the most natural cases.

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

1 Introduction and motivations

During the last decade, argumentation frameworks [35, 17] have evolved as a successful approach to formalize commonsense reasoning and decision making in multiagent systems (MAS). Application areas include issues such as joint deliberation, persuasion, negotiation, knowledge distribution and conflict resolution (e.g. [10, 36, 37, 11, 31]), among many others. Particularly, recent research by Leila Amgoud [3, 4] has shown that argumentation provides a sound setting to model reasoning about coalition formation in MAS. The approach is based on using conflict and preference relationships among coalitions to determine which coalitions should be adopted by the agents. This is done according to a particular argumentation semantics.

Alternating-time Temporal Logic (ATL) [2] is a temporal logic which can be used for reasoning about the behavior and abilities of agents under various rationality assumptions [26, 27, 15]. In ATL the key construct has the form \( \langle A \rangle \phi \), which expresses that a coalition \( A \) of agents can enforce the formula \( \phi \). Under a model theoretic viewpoint, \( \langle A \rangle \phi \) holds whenever the agents in \( A \) have a winning strategy for ensuring that \( \phi \) holds (independently of the behavior of \( A \)'s opponents). However, this operator accounts only for the theoretical existence of such a strategy, not taking into account whether the coalition \( A \) can be actually formed. Indeed, in order to join a coalition, agents usually require some kind of incentive (e.g. sharing common goals, getting rewards, etc.), since usually forming a coalition does not come for...
free (fees have to be paid, communication costs may occur, etc.). Consequently, several possible coalition structures among agents may arise, from which the best ones should be adopted according to some rationally justifiable procedure.

In this paper we present an argumentative approach to extend ATL for modelling coalitions. We provide a formal extension of ATL, \textsc{CoalATL}, by including a new construct \(\langle A\rangle \phi\) which denotes that the group \(A\) of agents is able to build a coalition \(B\), \(A \cap B \neq \emptyset\), such that \(B\) can enforce \(\phi\). That is, it is assumed that agents in \(A\) work together and try to form a coalition \(B\). The actual computation of the coalition is modelled in terms of a given argumentation semantics \[20\] in the context of coalition formation \[3\]. In a second step, we enrich \textsc{CoalATL} with goals. We address the question why agents should cooperate. Goals refer to agents’ subjective incentive to join coalitions. We show that the model checking problem of our framework is an extension of the model checking procedure used in ATL and we prove that it is \(\Delta^P_2\)-complete for a natural class of argumentation semantics.

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 incorporate goals to \textsc{CoalATL}, turning to model checking in Section 6. Finally, in Sections 7 and 8 we discuss related and future work and conclude.

**About this paper.** This article is based on \[12\], \[13\], and on an invited paper, combining both papers just mentioned, presented in \[14\]. Sections 2-4 and 6-8 are slight variations of \[12\]. However, some minor conceptual issues have been modified (in accordance with \[13\]) or added; discussions on future and related work has also been updated. Section 5 presents the work given in \[13\]. In this article, we substantially elaborated the work on the computational complexity of the model checking problem presented in Section 6. We included full proofs, especially about the \(\Delta^P_2\)-hardness, and added some additional complexity results.

## 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 (LATL)\[2\]** Let \(\text{agt} = \{a_1, \ldots, a_k\}\) be a nonempty finite set of all agents, and \(\Pi\) 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 \(\text{agt}\). The language \(L_{\text{ATL}}(\text{agt}, \Pi)\) is defined by the following grammar: \(\varphi ::= p \mid \neg \varphi \mid \varphi \land \varphi \mid \langle A\rangle \gamma\) where \(\gamma ::= \varphi \mid \neg \gamma \mid \gamma \land \gamma \mid \Pi \varphi \mid \Box \gamma \mid \bigtriangledown \gamma\). \(A \in \text{agt}\), and \(p \in \Pi\).

Formulae \(\varphi\) (resp. \(\gamma\)) are called state (resp. path) formulae.

Informally, \(\langle A\rangle \gamma\) expresses that agents \(A\) have a collective strategy to enforce \(\gamma\). ATL formulae include the usual temporal operators: \(\bigtriangledown\) (“in the next state”), \(\Box\) (“always from now on”) and \(\Pi\) (strict “until”). Additionally, \(\bigtriangledown\) (“now or sometime in the future”) can be defined as \(\bigtriangledown \varphi \equiv \top \cup \varphi\).

The semantics of ATL is defined by concurrent game structures.

**Definition 2 (CGS)\[2\]** A concurrent game structure (CGS) is a tuple \(\mathcal{M} = (\text{agt}, Q, \Pi, \pi, \text{Act}, d, o)\), consisting of: a set \(\text{agt} = \{a_1, \ldots, a_k\}\) of agents; set \(Q\) of states; set \(\Pi\) of atomic propositions; valuation of propositions \(\pi : Q \rightarrow 2^{\Pi}\); set \(\text{Act}\) of actions. Function \(d : \text{agt} \times Q \rightarrow 2^{\text{Act}}\) indicates the actions available to agent \(a \in \text{agt}\) 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 \(\alpha = \langle a_1, \ldots, a_k\rangle \in d(q)\) to another state \(q' = o(q, \alpha)\).

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 \(\lambda\). The set of all paths starting in \(q\) is defined by \(\Lambda_{\text{gt}}(q)\).

A (memoryless) strategy of agent \(a\) is a function \(s_a : Q \rightarrow \text{Act}\) such that \(s_a(q) \in d_a(q)\) We denote the set of such functions by \(\Sigma_a\). A collective strategy \(s_A\) for team \(A \subseteq \text{agt}\) 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 \(\Sigma := \Sigma_{\text{gt}}\).
The outcome of strategy $s_A$ in state $q$ is defined as the set of all paths that may result from executing $s_A$: \( \text{out}(q,s_A) = \{ \lambda \in \Lambda_{\text{out}}(q) \mid \forall i \in \mathbb{N}_0 \exists a = (a_1, \ldots, a_k) \in d(\lambda[i]) \forall a \in A \ (\alpha_a = s^a_A(\lambda[i]) \land \sigma(\lambda[i], a) = \lambda[i+1]) \} \), where $s^a_A$ denotes agent $a$’s part of the collective strategy $s_A$.

The semantics of $\text{ATL}$ is as follows.

**Definition 3 (atl Semantics [2])** Let a cgs $\mathcal{M} = \langle \text{Agt}, Q, \Pi, \pi, \text{Act}, d, \sigma \rangle$ and $q \in Q$ be given. The semantics of state formulae 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 $\mathcal{M}, q \not\models \varphi$
- $\mathcal{M}, q \models \varphi \land \psi$ iff $\mathcal{M}, q \models \varphi$ and $\mathcal{M}, q \models \psi$
- $\mathcal{M}, q \models \llbracket A \rrbracket \gamma$ iff there is $s_A \in \Sigma_A$ such that $\mathcal{M}, \lambda \models \gamma$ for all $\lambda \in \text{out}(q,s_A)$
- $\mathcal{M}, \lambda \models \varphi$ iff $\mathcal{M}, \lambda[0] \models \varphi$; and for path formulae by
- $\mathcal{M}, \lambda \models \Box \varphi$ iff $\mathcal{M}, \lambda[i] \models \varphi$ for all $i \in \mathbb{N}_0$
- $\mathcal{M}, \lambda \models \varphi \mathcal{U} \psi$ iff there is an $i \in \mathbb{N}_0$ with $\mathcal{M}, \lambda[i] \models \psi$, and $\mathcal{M}, \lambda[j] \models \varphi$ for all $0 \leq j < i$.

**3 Coalitions and Argumentation**

In this section we provide an argument-based characterization of coalition formation that will be used later to extend $\text{ATL}$. 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 [20], extended with a preference relation. The basic notion is that of a coalitional framework, which contains a set of elements $\mathcal{E}$ (usually seen as agents or coalitions), an attack relation (for modelling conflicts among elements of $\mathcal{E}$), and a preference relation between elements of $\mathcal{E}$ (to describe favorite agents/coalitions).

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

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

The set $\mathcal{E}$ in Definition 4 is intentionally generic, accounting for various possible alternatives. One alternative is to consider $\mathcal{E}$ as a set of agents $\text{Agt} = \{a_1, \ldots, a_k\}$: $\mathcal{CF} = (\mathcal{E}, A, \prec) \in \mathcal{CF}(\text{Agt})$. Then, a coalition is given by $C = \{a_1, \ldots, a_k\} \subseteq \mathcal{E}$ and “agent” can be used as an intuitive reference to elements of $\mathcal{E}$. Another alternative is to use a coalitional framework $\mathcal{CF} = (\mathcal{E}, A, \prec)$ based on $\mathcal{CF}(\text{Agt})$. Now elements of $\mathcal{E} \subseteq \mathcal{P}(\text{Agt})$ are groups or coalitions (where we consider singletons as groups too) of agents. Under this interpretation a coalition $C \subseteq \mathcal{E}$ is a set of sets of agents. Although “coalition” is already used for $C \subseteq \mathcal{E}$, we also use the intuitive reading “coalition” or “group” to address elements in $\mathcal{E}$.

Yet another way is not to use the specific structure for elements in $\mathcal{E}$, 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 $\mathcal{E}$ as a set of agents.
Example 1 Consider the following two coalitional frameworks: (i) $\mathcal{CF}_1 = (\mathcal{C}, \mathcal{A}, \preceq)$ where $\mathcal{C} = \{a_1, a_2, a_3\}$, $\mathcal{A} = \{(a_3, a_2), (a_2, a_1), (a_1, a_3)\}$ and agent $a_3$ is preferred over $a_1$, i.e. $a_1 \prec a_3$; and (ii) $\mathcal{CF}_2 = (\mathcal{C}', \mathcal{A}', \prec')$ where $\mathcal{C}' = \{(a_1), \{a_2\}, \{a_3\}\}$, $\mathcal{A}' = \{\{(a_3), \{a_2\}, \{a_1\}, \{a_1, a_3\}\}$ and group $\{a_3\}$ is preferred over $\{a_1\}$, i.e. $\{a_1\} \prec' \{a_3\}$. They capture the same scenario and are isomorphic but $\mathcal{CF}_1 \notin \mathcal{CF}(\{a_1, a_2, a_3\})$ and $\mathcal{CF}_2 \notin \mathcal{CF}(\{a_1, a_2, a_3\})$; 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 $\mathcal{CF} = (\mathcal{C}, \mathcal{A}, \preceq)$ be a coalitional framework. For $C, C' \in \mathcal{C}$, we say that $C$ attacks $C'$ iff $\mathcal{CAF}$. The attack relation represents conflicts between elements of $\mathcal{C}$; for instance, two agents may rely on the same (unique) resource or they may have disagreeing goals, which prevent them from cooperation. However, the notion of attack may not be sufficient for modelling conflicts, as some elements (resp. coalitions) in $\mathcal{C}$ may be preferred over others. This leads to the notion of defater which combines the notions of attack and preference.

Definition 5 (Defeater) Let $\mathcal{CF} = (\mathcal{C}, \mathcal{A}, \preceq)$ be a coalitional framework and let $C, C' \in \mathcal{C}$. 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 defater for $C'$.

Attacks and defeats are defined between single elements of $\mathcal{C}$. 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 $\mathcal{CF} = (\mathcal{C}, \mathcal{A}, \preceq)$ be a coalitional framework and $C \in \mathcal{C}$. We say that $C$ defends itself, if it is preferred over all attackers $C'$ of it.

Furthermore, $C$ is defended by a set $\mathcal{S} \subseteq \mathcal{C}$ of elements of $\mathcal{C}$ if, and only if, for all $C'$ defeating $C$ there is a coalition $C'' \in \mathcal{S}$ defeating $C'$.

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

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 \cite{dung1995argumentation, dung2000extensions}. 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 $\mathcal{CF} = (\mathcal{C}, \mathcal{A}, \preceq)$ is a (isomorphism invariant) mapping $\text{sem}$ which assigns to a given coalitional framework $\mathcal{CF} = (\mathcal{C}, \mathcal{A}, \preceq)$ a set of subsets of $\mathcal{C}$, i.e., $\text{sem}(\mathcal{CF}) \subseteq \mathcal{P}(\mathcal{C})$.\(^2\)

\(^2\)In \cite{dung2000extensions} these notions are defined the other way around, resulting in a different characterization of stable semantics.
Let $\mathcal{CF} = (\mathcal{E}, \mathcal{A}, \prec)$ be a coalitional framework. To formally characterize different semantics we define a function $F_{\mathcal{CF}} : \mathcal{P}(\mathcal{E}) \to \mathcal{P}(\mathcal{E})$ which assigns to a set of coalitions $\mathcal{S} \subseteq \mathcal{E}$ the coalitions defended by $\mathcal{S}$.

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

$$F_{\mathcal{CF}} : \mathcal{P}(\mathcal{E}) \to \mathcal{P}(\mathcal{E})$$

$$F_{\mathcal{CF}}(\mathcal{S}) = \{ C \in \mathcal{E} \mid C \text{ is defended by } \mathcal{S} \}$$

is called characteristic function\(^1\).

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

Example 2 Consider again the coalitional framework $\mathcal{CF}_1$ given in Example\(^2\). The characteristic function applied on the empty set results in $\{a_3\}$ since the agent is undefeated, $F(\emptyset) = \{a_3\}$. Applying $F$ on $F(\emptyset)$ 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 $F$.

We now introduce the first concrete semantics called coalition structure semantics, which was originally defined in\(^3\).

Definition 10 (Coalition structure $\text{sem}_{cs}$) Let $\mathcal{CF} = (\mathcal{E}, \mathcal{A}, \prec)$ be a coalitional framework. Then

$$\text{sem}_{cs}(\mathcal{CF}) := \bigcup_{i=1}^{\infty} F_{\mathcal{CF}}^i(\emptyset)$$

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

For a coalitional framework $\mathcal{CF} = (\mathcal{E}, \mathcal{A}, \prec)$ with a finite set $\mathcal{E}$, the characteristic function $F$ is continuous\(^4\) (according to Knaster-Tarski). We have the following straightforward properties of coalition structure semantics.

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

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 $F_{\mathcal{CF}_1}$ the coalitional framework $\mathcal{CF}_1$ has $\{a_1, a_3\}$ as coalitional structure.

(b) $\mathcal{CF}_3 := (\mathcal{E}, \mathcal{A}, \prec) \in \mathcal{CF} (\{a_1, a_2, a_3\})$ (shown in Figure\(^7\)b), is a coalitional framework with $\mathcal{E} = \{a_1, a_2, a_3\}$, $\mathcal{A} = \{a_1, a_2\}, (a_1, a_3), (a_2, a_1), (a_2, a_3), (a_3, a_1)\}$ and $a_4$ is preferred over $a_2, a_2 \prec a_3$, has the empty coalition as associated coalition str., i.e. $\text{sem}_{cs}(\mathcal{CF}) = \{\emptyset\}$.

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\(^20\) and can be used as a criterion for coalition formation (cf.\(^3\)).

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

---

\(^1\)We omit the subscript $\mathcal{CF}$ if it is clear from context.

\(^2\)Actually, it is enough to assume that $\mathcal{CF}$ is finitary (cf.\(^20\) Def. 27)).
(a) admissible extension iff $\mathcal{S}$ is conflict-free and $\mathcal{S}$ defends all its elements, i.e. $\mathcal{S} \subseteq \mathcal{F}(\mathcal{S})$.

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

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

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

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

Let $\text{sem}_{cs}$ (resp. $\text{sem}_{complete}$, $\text{sem}_{grounded}$, $\text{sem}_{preferred}$ and $\text{sem}_{stable}$) 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 1, the following holds:

$$\begin{align*}
\text{sem}_{cs}(\mathcal{CF}) &= \{\emptyset\} \\
\text{sem}_{admissible}(\mathcal{CF}) &= \{\{a_1\}, \{a_2\}, \{a_3\}, \{a_2, a_3\}\}
\end{align*}$$

Analogously, for the coalitional framework $\mathcal{CF}_1$ from Example 2, 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 ($\text{CoalATL}$). This logic extends ATL by new operators $\langle A \rangle$ for each subset $A \subseteq \text{Agt}$ of agents. These new modalities combine, or rather integrate, coalition formation into the original ATL cooperation modalities ($\langle A \rangle \varphi$). The intended reading of $\langle A \rangle \varphi$ is that the group $A$ of agents is able to form a coalition $B \subseteq \text{Agt}$ such that some agents of $A$ are also members of $B$, $A \cap B \neq \emptyset$, and $B$ can enforce $\varphi$. Coalition formation is modelled 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 $\langle A \rangle \varphi$). The new modality $\langle A \rangle$ provides a rather subjective view of the agents in $A$ and their power to create a group $B$, $A \cap B \neq \emptyset$, which in turn is used to reason about the ability to enforce a given property.

The language of $\text{CoalATL}$ is as follows.

Definition 12 ($\mathcal{L}_{\text{ATL'}}$) Let $\text{Agt} = \{a_1, \ldots, a_k\}$ be a finite, nonempty set of agents, and $\Pi$ 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 $\text{Agt}$. The language $\mathcal{L}_{\text{ATL'}}(\text{Agt}, \Pi)$ is defined by the following grammar:

$$\varphi ::= p \mid \neg \varphi \mid \varphi \land \varphi \mid \langle A \rangle \varphi \mid \langle A \rangle \Box \varphi \mid \langle A \rangle \varphi U \varphi \mid \langle A \rangle \varphi \mid \langle A \rangle \Box \varphi \mid \langle A \rangle \varphi U \varphi$$

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. Moreover, we assume that coalitional frameworks depend on the agents. Two initial groups of agents may have different skills to form coalitions. Consider for instance the following example.
Example 5 Imagine the two agents $a_1$ and $a_2$ are not able (because they do not have the money) to convince $a_3$ and $a_4$ to join. But $a_1$, $a_2$, and $a_3$ together have the money and all four can enforce a property $\varphi$. So $\{a_1, a_2\}$ are not able to build a greater coalition to enforce $\varphi$; but $\{a_1, a_2, a_3\}$ are. So we are not looking at coalitions per se, but how they evolve from others.

We assume that the argumentative semantics is the same for all states.

**Definition 13 (CGM)** A coalitional game model (CGM) is given by a tuple

$$\mathcal{M} = (\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem})$$

where $\langle \text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem} \rangle$ is a CGS, $\zeta : \mathcal{P}(\text{Agt}) \to (Q \to \mathcal{CF}(\text{Agt}))$ is a function which assigns a coalitional framework over $\text{Agt}$ to each state of the model subjective to a given group of agents, and $\text{sem}$ is an (argumentative) semantics as defined in Definition 8. The set of all such models is given by $\mathcal{M}(Q, \text{Agt}, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem})$.

A model provides an argumentation semantics $\text{sem}$ 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 not disjunct with the predetermined starting coalition. This leads to the notion valid coalition.

**Definition 14 (Valid coalition)** Let $A, B \subseteq \text{Agt}$ be groups of agents, $\mathcal{M} = (\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem})$ be a CGM and $q \in Q$.

We say that $B$ is a valid coalition with respect to $A, q,$ and $\mathcal{M}$ whenever $B \in \text{sem}(\zeta(A)(q)))$ and if $A \neq \emptyset$ then $A \cap B \neq \emptyset$. Furthermore, we use $\text{vc}_{\text{sem}}(A, q)$ 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.

**Remark 2** Note, that in [12] we assume that the members of the initial group $A$ work together, whatever the reasons might be. So group $A$ was added to the semantics. This ensured that agents in $A$ can enforce $\psi$ on their own, if they are able to do so. Even if $A$ is not accepted originally by the argumentation semantics, i.e. $A \not\in \text{sem}(\zeta(A)(q))$. Here, we drop this requirement. As pointed out in [12] the “old” semantics is just a special case of this new one: The operator from [12] can be defined as $\{A\}\gamma \lor \{A\}\gamma$.

Moreover, we changed the condition that the predefined group given in the coalitional operator must be a subset of the formed coalition, $A \subseteq B$, to the requirement that some member of the initial coalition (if $A \neq \emptyset$) should be in the new one, $A \cap B \neq \emptyset$. Both definitions make sense in different scenarios; however, the new one seems to be more generic.

The semantics of the new modality is given by

**Definition 15 (Coalition ATL Semantics)** Let a CGM $\mathcal{M} = (\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem})$ a group of agents $A \subseteq \text{Agt}$, and $q \in Q$ be given. The semantics of Coalitional ATL extends that of ATL, given in Definition 8, by the following rule ($\{A\}\psi \in \mathcal{L}_{\text{ATL}}(\text{Agt}, \Pi)$):

$$\mathcal{M}, q \models \{A\}\psi \iff \text{there is a coalition } B \in \text{vc}(A, q) \text{ such that } \mathcal{M}, q \models \langle B \rangle \psi.$$  

**Remark 3 (Different Semantics, $\models_{\text{sem}}$)** We have just defined a whole class of semantic rules for modality $\{ \cdot \}$. The actual instantiation of the semantics $\text{sem}$, for example $\text{sem}_{\text{stable}}, \text{sem}_{\text{pref}},$ and $\text{sem}_{\text{ex}}$ defined in Section 8 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 CGM $\mathcal{M}$ with an argumentation semantics $\text{sem}$ we write $\models_{\text{sem}}$ instead of $\models$.

The underlying idea of the semantic definition of $\{A\}\psi$ is as follows. A given (initial) group of agents $A \subseteq \text{Agt}$ is able to form a valid coalition $B$ (where $A$ and $B$ must not be disjoint), with respect to a given coalitional framework $\mathcal{CF}$ and a particular semantics $\text{sem}$, such that $B$ can enforce $\psi$.

Similarly to the alternatives to our definition of valid coalitions there are other sensible semantics for Coalition ATL. 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.
Figure 2: A simple CGM defined in Example 5.

Proposition 4 ([12]) Let $A \subseteq \text{Agt}$ and $\langle A \rangle \psi \in \mathcal{L}_{\text{ATL}}(\text{Agt}, \Pi)$. Then $\langle A \rangle \psi \rightarrow \bigvee_{B \subseteq \text{Agt}} (A \subseteq B) \psi$ is valid with respect to CGM's.

Compared to ATL, a formula like $\langle A \rangle \varphi$ does not refer to the ability of $A$ to enforce $\varphi$, but rather to the ability of $A$ to constitute a coalition $B$, such that $A \cap B \neq \emptyset$, and then, in a second step, to the ability of $B$ to enforce $\varphi$. Thus, two different notions of ability are captured in these new modalities. For instance, $\langle A \rangle \psi \land \neg \langle \emptyset \rangle \psi$ expresses that group $A$ of agents can enforce $\psi$, but there is no reasonable coalition which can enforce $\psi$ (particularly not $A$, although they possess the theoretical power to do so).

Example 6 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 $x \neq r$. Each agent can choose to perform action $\alpha$ or $\beta$. Action profiles and their outcomes are shown in Figure 3. The $*$ is used as a placeholder for any of the two actions, i.e. $* \in \{\alpha, \beta\}$. For instance, the profile $(\beta, \beta, *)$ leads to state $q_2$ whenever agent $a_1$ and $a_2$ perform action $\beta$ and $a_3$ either does $\alpha$ or $\beta$.

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 coaltional game $\mathcal{CGM} = (\mathcal{E}, \mathcal{A}, \prec)$ given in Example 3(b), that is, $\mathcal{E} = \text{Agf}$, $\mathcal{A} = \{(a_1, a_2), (a_1, a_3), (a_2, a_1), (a_2, a_3), (a_3, a_1)\}$ and $\mathcal{A}_2 = \{a_3\}$.

We formalize the example as the CGM $\mathcal{M} = (\text{Agf}, Q, \Pi, \pi, \text{Act}, \alpha, \zeta, \text{sem})$ where $\text{Agf} = \{a_1, a_2, a_3\}, Q = \{q_0, q_1, q_2, q_3\}$, $\Pi = \{r, s, t\}$, and $\zeta(q) = \mathcal{CGM}$ for all states $q \in Q$ and groups $A \subseteq \text{Agf}$. Transitions and the state labeling can be seen in Figure 3. Furthermore, we do not specify a concrete semantics $\text{sem}$ 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 $a_3$: $\mathcal{M}, q_0 \models \langle a_1, a_2 \rangle \neg r$. Indeed, $\{a_1, a_2\}$ and the grand coalition $\text{Agf}$ (since it contains $\{a_1, a_2\}$) are the only coalitions which are able to enforce $\neg r$; we have

$$\mathcal{M}, q_0 \models \neg \langle X \rangle \circ r$$

for all $X \subseteq \text{Agf}$ and $X \neq \{a_1, a_2\}$. Outcomes $s$ or $t$ can be enforced by $a_2$: $\mathcal{M}, q_0 \models \langle a_2 \rangle \circ (s \lor t)$. Agents $a_2$ and $a_3$ also have the ability to enforce a situation which agrees with both of their desired outcomes: $\mathcal{M}, q_0 \models \langle a_2, a_3 \rangle \circ (s \land t)$.

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

$$\mathcal{M}, q_0 \models \langle a_1, a_2 \rangle \circ r \land \neg \langle a_1 \rangle \circ r \land \neg \langle a_2 \rangle \circ r$$

for any semantics $\text{sem}$ introduced in Definition 5 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 $r$ (in $q_0$) because of [7]. 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 coaltional framework $\mathcal{CGM}$).
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 7** \textsc{coalatl} can be used to determine whether a coalition for enforcing a specific property exists. Assume that \textit{sem} represents the grounded semantics. For instance, the statement

\[ \mathcal{M}, q_0 \models_{\text{sem \text{grounded}}} \langle \emptyset \rangle \circ \top \]

expresses that there is a grounded coalition (i.e., a coalition w.r.t. the grounded semantics) which can enforce \( \circ \top \), namely the coalition \( \{a_2, a_3\} \). This result does not hold for all semantics; for instance, we have

\[ \mathcal{M}, q_0 \not\models_{\text{set-sem \text{grounded}}} \langle \emptyset \rangle \circ \top \]

with respect to the coalition structure semantics, since the coalition structure is the empty coalition and \( \mathcal{M}, q_0 \not\models \langle \emptyset \rangle \circ \top \).

In the following section we sketch of the language can be extended by an update mechanism, in order to compare different argumentative semantics using formulae inside the object language.

### 4.1 An update mechanism

In Example 7 we have shown that the underlying semantics of the coalition framework is crucial for the truth of a formula. We showed, for instance, that \( \langle \emptyset \rangle \circ \top \) is true w.r.t. the grounded semantics but false regarding the coalition structure semantics. This comparison took place on the meta-level; two \textit{cgms}' were defined, using grounded and coalition structure semantics, respectively. In this section, we introduce semantics as \textit{first-class citizens} in the object language. Therefore, we extend the language by \textit{semantic terms}, out of a set \( \Omega \), and an update operator \((\text{set-sem})\). Semantically, a \textit{cgm} \( \mathcal{M} \) is enriched by a \textit{denotation function} \([\cdot] : \Omega \rightarrow (\text{CF}(\text{Agt}) \rightarrow \mathcal{P}(\text{P}(\text{Agt})))\) which maps semantic terms to an argumentation semantics. The idea is that \((\text{set-sem}) \textit{sem})\) resets the semantics in \( \mathcal{M} \) to \([\text{sem}] \), where \( \text{sem} \in \Omega \). The intended reading for \((\text{set-sem}) \textit{sem})\) \( \varphi \) is that \( \varphi \) holds if the argumentation semantics is given by \([\text{sem}] \). We formally define the new language and its models.

**Definition 16** (\( \mathcal{L}_{\text{ATL}} \) plus update) As in Definition 12 let \( \text{Agt} = \{a_1, \ldots, a_k\} \) and \( \Pi \) be given. Let \( \Omega \) be a non-empty set, its elements are called semantic symbols (with typical element \( \text{sem} \)).

The logic \( \mathcal{L}_{\text{ATL}+\text{c}}(\text{Agt}, \Pi, \Omega) \) is given by all formulas of \( \mathcal{L}_{\text{ATL}}(\text{Agt}, \Pi) \) and for all \( \varphi \in \mathcal{L}_{\text{ATL}+\text{c}}(\text{Agt}, \Pi, \Omega) \) we also have \((\text{set-sem}) \textit{sem})\) \( \varphi \in \mathcal{L}_{\text{ATL}+\text{c}} \), where \( \text{sem} \in \Omega \).

**Remark 5** (Standard semantic terms) We assume that for all semantics defined in Definitions 8 and 11 there is a corresponding semantics term in \( \Omega \). For example, for the grounded semantics \( \text{sem}_{\text{grounded}} \) there is a term \( \text{sem}_{\text{grounded}} \) in \( \Omega \).

We need to define the denotation of the new syntactic objects.

**Definition 17** (\( \text{cgm} + \text{update} \)) A coalitional concurrent game structure with update \((\text{ccgs+u})\) is given by a tuple

\[ \mathcal{M} = (\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \textit{sem}, \Omega, []) \]

where \((\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \textit{sem})\) is a \( \text{cgm} \), \( \Omega \) is a non-empty set of semantic terms, and \([\cdot] : \Omega \rightarrow (\text{CF}(\text{Agt}) \rightarrow \mathcal{P}(\text{P}(\text{Agt})))\) is a denotation function, such that \([\text{sem}] \) is an argumentation semantics over \( \text{CF}(\text{Agt}) \) (cf. Definition 5) for all \( \text{sem} \in \Omega \).

In accordance with Remark 5 we assume that the denotation of semantic terms belonging to one of the “standard” semantics connects the terms with their semantics. That is, we assume, for instance, that the denotation of \( \text{sem}_{\text{grounded}} \) is \( \text{sem}_{\text{grounded}} \).

In addition to all semantic rules given before, we also need to interpret \((\text{set-sem})\).
Definition 18 (Semantics) Let $\mathcal{M} = \langle \text{Ag}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem}, \Omega, \cdot \rangle$ be a CCGS+u and $\text{sem} \in \Omega$. The semantics of $\text{CoalATL}$ plus update extends that of $\text{CoalATL}$, given in Definition 15, by the following rule ($\psi \in \mathcal{L}_{\text{ATL}^{++}}(\text{Ag}, \Pi, \Omega)$):

$$\mathcal{M}, q \models (\text{set-sem} \ \text{sem})\psi \iff \mathcal{M}^{\text{sem}}, \Pi, \text{Act}, d, o, \zeta, \Omega \models \psi$$

where $\mathcal{M}^{\text{sem}}$ is a CCGS+u equal to $\mathcal{M}$ but its argumentation semantics is given by (set to) $\text{[sem]}$.

Example 8 Let $\mathcal{M}'$ be the CCGS+u which corresponds to the CGM $\mathcal{M}$ given in Example 7 extended by a set $\Omega$ of semantic terms and denotation function $\text{[sem]}$. We can state the relation between grounded and coalition structure semantics directly on the object level:

$$\mathcal{M}, q_0 \models (\text{set-sem} \ \text{sem}_{\text{grounded}})\Omega \odot t \land \neg (\text{set-sem} \ \text{sem}_{\text{cs}})\emptyset \odot t.$$ 

5 Cooperation and Goals

Why should agents join coalitions? They must have reasons to do so. Here, we consider goals as the driving force, and consequently, we assume that agents act to reach their goals. Firstly, we propose an abstract goal framework. Secondly, we use specific languages for goals and objectives, and we propose ATL as a suitable language to capture agents’ goals. Finally, we implement goals into the semantics of $\text{CoalATL}$, discuss its benefits and illustrate it with an example.

5.1 Goals and Agents

Pro-activeness and social ability are among the widely accepted characteristics of intelligent agents [11]. In BDI frameworks, goals (or desires) and beliefs play an important role [10, 38].

We believe that also the social ability to join coalitions, should be based on some incentive. Agents are usually not developed to offer their services for free. Also in the agent programming community several types of goals (e.g. achievement or maintenance goals) are commonly considered as an agent’s driving force. Here, we present a simple abstract framework to deal with these notions.

Definition 19 ($\mathcal{G}$, goal mapping $g$) Let $\mathcal{G}_a$ be a non-empty set of elements (set of goals), one for each agent $a \in \text{Ag}$, and $\mathcal{G} := \bigcup_{a \in \text{Ag}} \mathcal{G}_a$. By “$g$” we denote a typical element from $\mathcal{G}$. A goal mapping is a function $g : \text{Ag} \rightarrow (Q^+ \rightarrow \mathcal{P}(\mathcal{G}))$ assigning a set of goals to a given sequence of states and agent.

So, a goal mapping assigns a set of goals to a history, depending on an agent. This is needed to introduce goals into CGM’s. The history dependency can be used, for instance, to model when a goal should be removed from the list: An agent having a goal $\Diamond s$ may drop it after reaching a state in which $s$ holds. Alternatively, a model update mechanism can be used to achieve the same regarding state-based goal mappings; however, in our opinion the former is more elegant.

An agent might have several goals. Often, goals can not be reached simultaneously which requires means to decide which goal should be selected first. We model this by a preference ordering.

Definition 20 (Goal preference relation) A goal preference ordering ($gp$-ordering) $\preceq$ over a set of goals $\mathcal{G}' \subseteq \mathcal{G}$ is a complete, transitive, antisymmetric, and irreflexive relation $\preceq \subseteq \mathcal{G}' \times \mathcal{G}'$. We say that a goal $g$ is preferred to $g'$ if $g \preceq g'$.

Given a goal mapping $g_a$ for $a \in \text{Ag}$ we assume that there implicitly also is a $gp$-ordering $\preceq_a$ (a’s $gp$-ordering).

So far, we did not say how goals can be actually used to form coalitions. We assume, given some task, that agents having goals satisfied or partly satisfied by the outcome of the task are willing to cooperate to bring about the task. In the following we will use the notion objective (or objective formula) to refer to both the task itself and the outcome of it. A typical objective is written as $o$. Agents which have goals fulfilled or at least partly supported by objective $o$ are possible candidates to participate in a coalition aiming at $o$.

We say that an objective $o$ satisfies goal $g$, $o \prec g$, if the complete goal $g$ is fulfilled after $o$ has been accomplished. If a goal is (partly) satisfied by $o$ we say that $o$ supports $g$, $o \prec^* g$; i.e. there is another goal $g'$ which is a subgoal of $g$ and which is satisfied. These notions will be made precise in the following sections. Intuitively, an objective $\Box t$ satisfies goal $\Box (t \lor s)$ an supports goal $\Diamond t$. 

Inteligencia Artificial 46(2010)
5.2 Specifying Goals and Objectives

In this section, we propose to use ATL-path formulas for specifying goals. It has been shown that temporal logics like linear-time temporal logic (LTL) and computation tree logic (CTL) can be used as goal specification languages \(^3\)\(^5\).\(^7\).

Goals formulated in LTL are very intuitive. Formulae like \(\Diamond \text{rich} \) (eventually being rich), \(\Box \text{takeUmbrella} \) (take umbrella in the next moment), or \(\square \Diamond \text{sleep} \) (going to sleep again and again) have clear interpretations. But goals formulated in CTL can be ambiguous. A goal like \(A \Diamond \text{rich} \) does not seem fundamentally different from \(\Diamond \text{rich} \) from the agent’s point of view. Its goal of being rich in the future can be read implicitly as being rich in all possible futures; only one of them can actually become true and in that particular one the agent wants to be rich.

In this section we will use ATL for expressing agents’ goals. At first glance, this seems to contradict the statement made above since CTL can be seen as a special case (the one agent fragment) of ATL. But this is not the case: CTL refers to a purely temporal setting whereas ATL talks about abilities of agents. Here is a clarifying example. Assume that there are two agents \(a\) and \(b\) both having access to the same critical section; that is, either \(a\) or \(b\) should access this section but not both. In such a case it is reasonable that agent \(a\) has the goal of preventing \(b\) to enter this section on its own: \(\neg \langle b \rangle \Diamond \text{critical} \). However, it might be acceptable for \(a\) that \(b\) together with another agent \(c\) enters the critical section because then \(c\) has to unlock resources \(a\) could use instead. Let us consider a more detailed example.

Example 9 (ATL-goals) In the example we consider two agents \(a\) and \(b\). Both agents can perform actions \(\alpha\) and \(\beta\). The first agent, leader of a research group, would like to get a better salary \(\langle bs \rangle\) and wants to retain the power to decide when to take vacation \(\langle vac \rangle\). So, \(a\)’s goal can be expressed as \(\gamma \equiv \Box (\neg \langle \text{now} \rangle \rightarrow \langle bs \rangle \land \langle vac \rangle)\). Interpreting the models shown in Figure 3 purely temporally (i.e. without action profiles) the CTL formula \(E_3\) is satisfied in \(q_0\) in both models: There are \(q_0\)-paths which satisfy \(\gamma\). On the other hand, \(A \gamma\) is false in both models in \(q_0\).

Now agent \(b\) enters the scene. A higher salary would require \(a\) to move to a company in which the agent has a boss who might be able to decide on \(a\)’s vacation (depending on the contract). Actually, although \(a\) would like to have a better salary he prefers to decide on his vacation on his own. Thus, his goal can be reformulated to \(\gamma' \equiv \Box (\neg \langle \text{now} \rangle \rightarrow \langle bs \rangle \land \langle \text{now} \rangle \rightarrow \langle \text{vac} \rangle)\). In the first model \(b\) does not have the power to decide on \(a\)’s vacation but \(b\) has this ability in the second model.

This quite simplistic example shows that ATL formulae can make sense as goal specification language.

Definition 21 (ATL-Goal) Let \(\gamma, \gamma'\) be ATL path formulae. An ATL-goal has the form \(\gamma \lor \gamma'\)\(^6\).\(^6\).

Note that goals can easily be defined as Coalitional ATL formulae; however, due to simplicity we stick to pure ATL formulae.

It remains to define the objective language. Consider the Coalitional ATL formula \(\langle A \rangle \gamma\). The question is whether there is a rational group to bring about \(\gamma\); thus, only agents which gain advantage when \(\gamma\) is

---

\(^5\)The operator \(A\) refers to all possible paths starting in a state. In ATL this operator can be expressed as \(\langle\emptyset\rangle\).

\(^6\)Note that \(\gamma \land \gamma'\) is not an ATL path formula anymore.
fulfilled should cooperate. Hence, we consider $\gamma$ as objective; so, in general all Coalitional $\text{ATL}$ path formulae.

**Definition 22 (Coal$\text{ATL}$-objective)** An coal$\text{ATL}$-objective is an $L_{\text{ATL}^\gamma}$ path formula.

### 5.3 Goals as a Means for Cooperation

In this section we link together Coalitional $\text{ATL}$ with the goal framework described above. The syntax of the logic is given as in Section 4. The necessary change takes place in the semantics. We redefine what it means for a coalition to be valid.

Up to now valid coalitions were solely determined by coalitional frameworks. Conflicts represented by such frameworks are a coarse, but necessary, criterion for a successful coalition formation process. However, nothing is said about incentives to join coalitions, only why coalitions should not be joined.

Goals allow to capture the first issue. For a given objective formula $o$ and a finite sequence of states, called *history*, we do only consider agents which have some goal supported by the current objective. 

However, nothing is said about incentives to join coalitions, only why coalitions should not be joined.

**Definition 23 (CGM with goals)** A CGM with goals (CGM$^g$) $M$ is given by a model of $M(Q, \text{Agent}, II, \text{sem}, \zeta)$ extended by a set of goals $G$ and a goal mapping $g$ over $G$. The set of all such models is denoted $M^g(Q, \text{Agent}, II, \text{sem}, G, g)$ or just $M^g$ if we assume standard naming.

To define the semantics we need some additional notation. Given a path $\lambda \in Q^\omega$ we use $\lambda[i, j]$ to denote the sequence $\lambda[i] \lambda[i + 1] \ldots \lambda[j]$ for $i, j \in N_0 \cup \{\infty\}$ and $i < j$. A *history* is a finite sequence $h = q_1 \ldots q_n \in Q^+$, $h[i]$ denotes state $q_i$ if $n \geq i$, $q_n$ for $i \geq n$, and $\varepsilon$ for $i < 0$ where $i \in Z \cup \{\infty\}$. Furthermore, given a history $h$ and a path or history $\lambda$ the combined path/history starting with $h$ extended by $\lambda$ is denoted by $h \circ \lambda$.

Finally, we present the semantics of coal$\text{ATL}$ with goals. It is similar to Definition 15. Here, however, it is necessary to keep track of the steps (i.e. visited states) made to determine the goals of the agents.

**Definition 24 (Goal-based semantics of $L_{\text{ATL}^\gamma}^g$)** Let $M$ be a CGM$^g$, $q$ a state, $\varphi, \psi$ state-, $\gamma$ a path formula, and $i, j \in N_0$. Semantics to $L_{\text{ATL}^\gamma}^g$ formulae is given as follows:

- $M, q, \tau \models p$ iff $p \in \pi(q)$
- $M, q, \tau \models \varphi \land \psi$ iff $M, q, \tau \models \varphi$ and $M, q, \tau \models \psi$
- $M, q, \tau \models \neg \varphi$ iff not $M, q, \tau \models \varphi$
- $M, q, \tau \models [\langle A \rangle \gamma$ iff there is a strategy $s_A \in \Sigma_A$ such that for all $\lambda \in \text{out}(q, s_A)$ it holds that $M, \lambda, \tau \models \varphi$
- $M, q, \tau \models [\langle A \rangle \gamma$ iff there is $A' \in \text{vc}^g(q, A, \gamma, \tau)$ such that $M, q, \tau \models [\langle A \rangle \gamma$
- $M, \lambda, \tau \models \varphi \iff M, \lambda[0], \tau \models \varphi$
- $M, \lambda, \tau \models \Box \varphi$ iff for all $i$ it holds that $M, \lambda[i], \tau \circ \lambda[1, i] \models \varphi$
- $M, \lambda, \tau \models \Box \varphi$ iff it holds that $M, \lambda[1], \tau \circ \lambda[1] \models \varphi$
- $M, \lambda, \tau \models \varphi \cup \psi$ iff there is a $j$ such that $M, \lambda[j], \tau \circ \lambda[1, j] \models \psi$ and for all $0 \leq i < j$ it holds that $M, \lambda[i], \tau \circ \lambda[1, i] \models \varphi$.

Ultimately, we are interested in $M, q \models \varphi$ defined as $M, q, q \models \varphi$.

We have to define when a goal is satisfied. Although the definition of *support* can be defined similarly, we focus on the former notion only.

**Definition 25 (Satisfaction of goals)** Let $g$ be an $\text{ATL}$-goal, $o$ an $L_{\text{ATL}^\gamma}^g$-objective, and $\tau \in Q^+$. We say that objective $o$ satisfies $g$, for short $o \models_{M, \tau, B} g$, with respect to $M, \tau$, and $B$ if, and only if, there is a strategy $s_B \in \Sigma_B$ such that
1. for all $\lambda \in \text{out}(\tau[\infty], s_B)$ it holds that $\mathcal{M}, \lambda, \tau \models \rho$ implies $\mathcal{M}, \lambda \models g$, and

2. that there is some path $\lambda \in \text{out}(\tau[\infty], s_B)$ with $\mathcal{M}, \lambda, \tau \models o$.

A goal is satisfied by an objective if each path (enforceable by $B$) that satisfies the objective does also satisfy the goal. That is, satisfaction of the objective will guarantee that the goal becomes true. The second condition ensures that the coalition actually has a way to bring about the goal. We show later, however, that the second condition is superfluous using the semantics defined in Definition 24.

It remains to define the semantics for combined (by conjunction) ATL path formulae. Therefore, we extend the ordinary semantics of ATL (given in Definition 3) by the following semantic rule: $\mathcal{M}, \lambda, \tau \models \rho$ if, and only if, $\mathcal{M}, \lambda \models \rho$. The all the new functionality provided by goals is captured by the new valid coalition function $\text{vc}^g$

**Definition 26 (Valid coalitions, $\text{vc}^g(q, A, o, \tau)$)** Let $\mathcal{M} \in \mathcal{M}^g$, $\tau \in Q^+$, $A, B \subseteq \text{Agt}$, $o$ an $\text{CoalATL}$ objective.

We say that $B$ is a valid coalition after $\tau$ with respect to $A$, $o$, and $\mathcal{M}$ if, and only if,

1. $B \in \text{sem}(\zeta(A)(\tau[\infty]))$, $A \cap B \neq \emptyset$ if $A \neq \emptyset$, and

2. there are goals $g_b, g_{b_0}$, one per agent $b \in B$, such that $o \leftarrow_{\mathcal{M}, \tau, B} g_{b_0} \land \cdots \land g_{b|B|}$

The set $\text{vc}^g(q, A, o, \tau)$ consists of all such valid coalitions wrt to $\mathcal{M}$.

Thus, for the definition of valid coalitions among other things, a goal mapping, a function $\zeta$ and a sequence of states $\tau$ are required. The intuition of $\tau$ is that it represents the history (the sequence of states visited so far including the current state). So, $\tau$ is used to determine which goals of the agents are still active.

Coming back to the satisfaction of goals. Let $\langle A \rangle^\prime$ denote a coalitional operator with semantics equal to $\langle A \rangle$ but without the second condition in Definition 25 (note that this definition is needed to determine the valid coalitions).

**Proposition 6** Let $\gamma$ be an $\text{CoalATL}$ path formula, $\mathcal{M} \in \mathcal{M}^g$, $q \in Q_{\mathcal{M}}$, $A \subseteq \text{Agt}$, and $\tau \in Q^+$. Then we have that $\mathcal{M}, q, \tau \models \langle A \rangle^\prime \gamma$ if, and only if, $\mathcal{M}, q, \tau \models \langle A \rangle \gamma$.

**Proof.** Firstly, observe that $\forall \lambda \in \text{out}(q, s_B) (\lambda \models \gamma)$ implies $\exists \lambda \in \text{out}(q, s_B) (\lambda \models \gamma)$ since the outcome is never empty. Then we have $\mathcal{M}, q, \tau \models \langle A \rangle \gamma$ if $\exists B \in \text{vc}^g \exists B \in \Sigma_B \forall \lambda \in \text{out}(q, s_B) (\lambda, \tau \models \gamma)$ iff $\exists B \in \text{vc}^g \exists B \in \Sigma_B (\forall \lambda \in \text{out}(q, s_B) (\lambda, \tau \models \gamma)$ and $\exists \lambda \in \text{out}(q, s_B) (\lambda, \tau \models \gamma)) (\ast)$ if $\exists B \in \text{vc}^g (\forall g \in \gamma \models (B)g)$ where $\text{vc}^g$ is equal to $\text{vc}^g$ except for condition 2 in Definition 25. Finally, it is easy to observe that the equivalence ($\ast$) holds since we can just take $s_B$ as goal satisfaction strategy. We omitted the model and the parameters of $\text{vc}^g$ ad $\text{vc}^g$.

**Proposition 7** We have that $\mathcal{M}, q, \tau \models \langle A \rangle \gamma$ if there is a coalition $B \in \text{vc}(A, q)$ and goals $g_b, g_{b_0}(\tau)$ one for each $b \in B$ such that $\mathcal{M}, q, \tau \models \langle B \rangle (\gamma \land \bigwedge_{b \in B} g_b)$.

**Remark 8 (Preferred goals)** In the abstract goal framework presented in Section 5.1, we defined a preference ordering over goals. The GP-orderings highly influence the coalition formation process. However, for this paper we decided to focus on the pure goal framework since the interplay between the formation process becomes much more sophisticated if preferences are taken into account. We just give a brief motivation for preferences and why they increase the complexity of coalition building.

The set of valid coalitions consists of all coalitions which are acceptable/conflict-free (according to a coalitional framework) and in which all agents have an incentive to join the coalition (that is, some goal has to be satisfied/supported). Let us consider two valid groups $B$ and $B'$ both containing the agent $a$. Both groups are somewhat appealing for since they satisfy some of his goals, say $B$ (resp. $B'$) can bring about $g$ (resp. $g'$). In our framework $B$ and $B'$ are treated equally good. Is this reasonable? From an abstract level it is; however, a finer grained analysis should incorporate the preferences between goals. If, for instance, $g$ is preferred over $g'$ agent a should rather go for coalition $B$ instead of $B'$. The agent would prefer to bring about $g$ thus joining $B$. On the other hand, if $a$ refuses to join $B'$ it might be possible, by
a symmetric argument, that another agent, say b, refuses to take part in B, such that in the end neither B nor $B'$ will form. Of course, in such a situation both agents prefer to bring about their less preferred goals. This is still better than getting nothing.

This reasoning very much reminds on game theoretic rationality concepts. For example, the motivation behind a Nash equilibrium strategy shows a strong connection: No agent has an incentive to unilaterally choose another strategy. Even closer are concepts from cooperative game theory. This discussion shows how interesting the incorporation of a preference ordering over goals is. However, this is also the reason why we did not incorporate it in the formation process here, it would be out of the scope of this paper. Our current research deals with this issue.

5.4 Progression of ATL goals

A goal mapping takes the history into account to be able to reflect if a goal has become fulfilled. For example, if an agent has goal $\Diamond p$ and $p$ became satisfied in a state on the current history the goal should be marked as completed in the following state. (Of course, a new goal in this state can again be $\Diamond p$.) Another, more practical but also restricted option, is to consider an initial goal base $\mathcal{GB}$ and modify, specialize or remove, the formulae according to the steps taken. So, goal $\Diamond p \land \Box q$ should be specialized to $\Box q$ if a state is reached in which $p$ holds. In [6] such a progression procedure is presented for first-order LTL.

6 Model Checking ATL\( ^c \)

In this section we present an algorithm for model checking $\text{CoalATL}^c$ formulae. The model checking problem is given by the question whether a given $\text{CoalATL}^c$ formula follows from a given $\text{CGM}$ $\mathcal{M}$ and state $q$, i.e. whether $\mathcal{M}, q \models \varphi$ [18]. In [2] it is shown that model checking $\text{ATL}$ 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 $O(m \cdot l)$.

6.1 Easiness

For $\text{CoalATL}^c$ we also have to treat the new coalitional modalities in addition to the normal $\text{ATL}$ constructs. Let us consider the formula $\langle\langle A \rangle\rangle \psi$. According to the semantics of $\langle\langle A \rangle\rangle$, given in Definition 15 we must check whether there is a coalition $B$ such that (i) if $A \neq \emptyset$ then $A \cap B \neq \emptyset$, (ii) $B$ is acceptable by the argumentation semantics, and (iii) $\langle\langle B \rangle\rangle \psi$. The number of possible candidate coalitions $B$ which satisfy (i) and (ii) is bounded by $|\mathcal{P}(\mathcal{AGT})|$. 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.

Both considerations together suggest that the model checking complexity has two computationally hard parts: exponentially many calls to $\langle\langle B \rangle\rangle \psi$ and the computation of the argumentation semantics. Indeed, Theorem 10 will support this intuition. However, we show that it is possible to "combine" both computationally hard parts to obtain an algorithm which is in $\Delta^P_2 = \text{P}^{\text{NP}}$, if the computational complexity to determine whether a given coalition is acceptable is in $\text{NP}$.

For the rest of this section, we will denote by $\text{VER}_{\text{sem}}(\mathcal{CF}, A)$ the verification problem (cf. [22]) which represents the problem whether for a given argumentation semantics $\text{sem}$, a coalitional framework $\mathcal{CF}$, and coalition $A \subseteq \mathcal{AGT}$ we have that $A \in \text{sem}(\mathcal{CF})$. Given some complexity class $\mathcal{C}$, we use the notation $\text{VER}_{\text{sem}} \in \mathcal{C}$ to state that the verification problem with respect to the semantics $\text{sem}$ is in $\mathcal{C}$.

In [12] it is stated that $\text{VER}_{\text{sem}} \in P$ for all semantics introduced in Definition 11. Unfortunately, this result is incorrect for the preferred semantics. There is a flaw in the way maximal sets of coalitions are treated. Actually, from [20] it follows that $\text{VER}_{\text{sem}} \in P$ for $\text{sem} \in \{\text{semadmissible}, \text{semgrounded}, \text{semstable}\}$ and in [19] it was shown that $\text{VER}_{\text{sempreferred}} \in \text{coNP}$-complete. In the following proposition, we summarize these results and also treat the complete semantics. For an overview we refer to [24].

Proposition 9 ([20] [19]) We have that $\text{VER}_{\text{sem}} \in P$ for all semantics $\text{sem} \in \{\text{semadmissible}, \text{semgrounded}, \text{semstable}, \text{semcomplete}\}$ and $\text{VER}_{\text{sempreferred}} \in \text{coNP}$-complete.
function mcheck(_M, q, \varphi_);  
Given a CGM \(\mathcal{M} = \langle \text{Ag}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem} \rangle\), a state \(q \in Q\), and \(\varphi \in \mathcal{L}_{ATL}(\text{Ag}, \Pi)\) the algorithm returns \(\top\) if, and only if, \(\mathcal{M}, q \models_{\text{sem}} \varphi\).

\[
\text{case } \varphi \text{ contains no } \langle B \rangle \text{: if } q \in \text{mcheck}_{\text{ATL}}(\mathcal{M}, \varphi) \text{ return } \top \text{ else } \bot
\]

\[
\text{case } \varphi \text{ contains some } \langle B \rangle \text{:}
\]

\[
\begin{align*}
&\text{case } \varphi \equiv \neg \psi \text{: return } \neg (\mathcal{M}, q, \psi) \\
&\text{case } \varphi \equiv \psi \lor \psi' \text{: return } \text{mcheck}(\mathcal{M}, q, \psi) \lor \text{mcheck}(\mathcal{M}, q, \psi') \\
&\text{case } \varphi \equiv \langle A \rangle \top \psi \text{: Label all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi) = \top \text{ with a new proposition } \text{yes} \text{ and return } \text{mcheck}(\mathcal{M}, q, \langle A \rangle \top \text{yes}); T \text{ stands for } \Box \text{ or } \nonumber \\
&\text{case } \varphi \equiv \langle A \rangle \psi U \psi' \text{: Label all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi) = \top \text{ with a new proposition } \text{yes}_1, \text{all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi') = \top \text{ with a new proposition } \text{yes}_2 \text{ and return } \text{mcheck}(\mathcal{M}, q, \langle A \rangle \psi \text{yes}_1 U \text{yes}_2) \\
&\text{case } \varphi \equiv \langle A \rangle \psi \text{ and } \psi \text{ contains some } \langle C \rangle \text{: Label all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi) = \top \text{ with a new proposition } \text{yes} \text{ and return } \text{mcheck}(\mathcal{M}, q, \langle A \rangle \top \text{yes}); T \text{ stands for } \Box \text{ or } \nonumber \\
&\text{case } \varphi \equiv \langle A \rangle \psi U \psi' \text{ or } \psi' \text{ contain some } \langle C \rangle \text{: Label all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi) = \top \text{ with a new proposition } \text{yes}_1, \text{all states } q' \text{ where } \text{mcheck}(\mathcal{M}, q', \psi') = \top \text{ with a new proposition } \text{yes}_2 \text{ and return } \text{mcheck}(\mathcal{M}, q, \langle A \rangle \psi \text{yes}_1 U \text{yes}_2) \\
&\text{case } \varphi \equiv \langle A \rangle \psi \text{ and } \psi \text{ contains no } \langle C \rangle \text{: Non-deterministically choose } B \in \mathcal{P}(\text{Ag}) \text{ if}
\begin{align*}
&(1) B \in (\text{sem}(\langle A \rangle(q))), \\
&(2) B \neq \emptyset \text{ then } A \cap B \neq \emptyset, \text{ and} \\
&(3) q \in \text{mcheck}_{\text{ATL}}(\mathcal{M}, \langle B \rangle \psi) \\
&\text{then return } \top \text{ else } \bot
\end{align*}
\]

function mcheck_{\text{ATL}}(_M, \varphi_);

Given a CGS \(\mathcal{M} = \langle \text{Ag}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem} \rangle\) and \(\varphi \in \mathcal{L}_{ATL}(\text{Ag}, \Pi)\), the standard ATL model checking algorithm (cf. \[12\]) returns all states \(q\) with \(\mathcal{M}, q \models_{\text{ATL}} \varphi\).

\[
\text{\textbullet return } \{ q \in Q \mid \mathcal{M}, q \models_{\text{ATL}} \varphi \}
\]

Figure 4: A model checking algorithm for \text{COALATL}.

**Proof.** It remains to show the case for \(\forall \mathcal{E} \mathcal{R}_{\text{sem-complete}}\). Note that computing \(\mathcal{F}_{CF}(\mathcal{G})\) for a given \(\mathcal{G}\) can be done in polynomial time. Therefore, to check whether a set of coalitions \(\mathcal{G}\) is complete, we can check that it is admissible and that all elements of \(\mathcal{F}_{CF}(\mathcal{G})\) are already contained in \(\mathcal{G}\). Both checks can be done in polynomial time. Thus \(\forall \mathcal{E} \mathcal{R}_{\text{sem-complete}} \in \mathcal{P}\).  

In Figure 4 we propose a model checking algorithm for \text{COALATL}. The complexity result given in the next theorem is moduli the complexity needed to solve the verification problem \(\forall \mathcal{E} \mathcal{R}_{\text{sem}}\).

**Theorem 10 (Model checking \text{COALATL} [12])** Let a CGM \(\mathcal{M} = \langle \text{Ag}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem} \rangle\) be given, \(q \in Q\), \(\varphi \in \mathcal{L}_{ATL}(\text{Ag}, \Pi)\), and \(\forall \mathcal{E} \mathcal{R}_{\text{sem}} \in \mathcal{C}\). Model checking \text{COALATL} with respect to the argumentation semantics \text{sem}[^7] is in \(\mathcal{P}^{\text{NP}}\).

**Proof.** The algorithm \text{mcheck}_{\text{ATL}} is \(\mathcal{P}\)-complete \[2\]. The number of modalities \(\langle \cdot \rangle\) is bounded by \(|\varphi|\). Thus, the last case can only be performed a polynomial number of times (wrt. the length of \(\varphi\)). The complexity of the last case is as follows. Firstly, \(B\) is guessed and then verified. The verification is

[^7]: That is, whether \(\mathcal{M}, q \models_{\text{sem}} \varphi\).
performed by an oracle call (with complexity $\mathcal{O}$) to check whether $B \in \text{sem}(\zeta(A)(q))$ and two additional steps which can be performed by a deterministic Turing machine in polynomial time.

The last theorem gives an upper bound for model checking $\text{CoalATL}$ with respect to an arbitrary but fixed semantics $\text{sem}$. A finer grained classification of the computational complexity of $\forall \forall \forall_{\text{sem}}$ allows to improve the upper bound given in Theorem 10. Assume that $\forall \forall \forall_{\text{sem}} \in \text{NP}$; then, a witness can be non-deterministically guessed together with the coalition $B \in \mathcal{P}(\kappa t)$ and then, it is checked whether $B$ satisfies the three conditions (1-3) in ($\ast$). Each of the three cases can be done in deterministic polynomial time. Hence, the verification of $\mathcal{M}, q \models \langle A \rangle ^{\ast} \psi$, in the last case, meets the “guess and verify” principle which is characteristic for problems in $\text{NP}$. This brings the overall complexity of the algorithm to $\Delta^p_2$.

**Corollary 11 ([12])** If $\forall \forall \forall_{\text{sem}} \in \text{NP}$ then model checking $\text{CoalATL}$ is in $\Delta^p_2$ with respect to $\text{sem}$.  

**Proof.** Since $\forall \forall \forall_{\text{sem}}$ is in $\text{NP}$ there is a deterministic Turing Machine $M$ that runs in polynomial time $p(n)$ (where $n$ is the length of the input $(A, \mathcal{C}F)$) that accepts $(A, \mathcal{C}F, w)$ for some witness $w$ of length less or equal $p(n)$ iff $A \in \text{sem}(\mathcal{C}F)$ (otherwise it does not accept $(A, \mathcal{C}F, w)$ for all witnesses $w$ with $|w| \leq p(n)$). Now, in the model checking algorithm, we extend the non-deterministic guess of the coalition $B$ in ($\ast$) by also guessing a witness $w$ for the input of machine $M$. This can be implemented by a single non-deterministic machine (e.g., using the deterministic machine $M$ as oracle or implementing it directly). Then, the whole algorithm is in $\text{D}^{\text{NP}} \text{P} = \text{P}^{\text{NP}} = \Delta^p_2$. ■

In the line with Proposition 9 we modify the result from [12] as follows.

**Corollary 12** Model checking $\text{CoalATL}$ is in $\Delta^p_2$ for $\text{sem}_{\text{admissible}}$, $\text{sem}_{\text{complete}}$ and $\text{sem}_{\text{stable}}$.

The following result is immediate as $\forall \forall \forall_{\text{sem}_{\text{preferred}}}$ is $\text{coNP}$-complete.

**Corollary 13** Model checking $\text{CoalATL}$ is in $\Delta^p_3$ for $\text{sem}_{\text{preferred}}$.

As the next proposition shows, the model checking algorithm can also be improved in the cases that only polynomially many coalitions are acceptable wrt to the semantics and that all these coalitions can be computed in polynomial time.

**Proposition 14** Model checking $\text{CoalATL}$ is $\text{P}$-complete for semantics $\text{sem}$ that only accept polynomially many coalitions and for which it is possible to enumerate all theses coalitions in polynomial time with respect to the size of the model and the length of the formula.

**Proof.** For a formula $\langle A \rangle ^{\gamma}$ we verify whether $\langle B \rangle ^{\gamma}$ for all the polynomially many coalitions $B$ acceptable by $\text{sem}$. Completeness follows from the completeness of model checking pure $\text{ATL}$ [2]. ■

Since the grounded semantics is characterized be the smallest fixed point there only is a unique coalition. So, the following result is immediate.

**Corollary 15** Model checking $\text{CoalATL}$ is $\text{P}$-complete for $\text{sem}_{\text{grounded}}$.

### 6.2 Definitions and NP/coNP-hardness

In the next section we show that model checking $\text{CoalATL}$ is $\Delta^p_2$-hard even for the very simple argumentation semantics that can essentially characterize all truth assignments of Boolean formulae. In this section, we firstly show that the model checking problem is $\text{NP}$-hard and $\text{coNP}$-hard by reducing SAT [33] (satisfiability of Boolean formulae (in positive normal form)) to model checking $\text{CoalATL}$ and introduce the basic definitions needed for the $\Delta^p_2$-hardness proof presented in the following section.

Let $\varphi = \varphi(X)$ be a Boolean formula in positive normal form over the Boolean variables $X := \{x_1, \ldots, x_n\}$. A truth assignment of a Boolean formula $\varphi(X)$ is a mapping $X \rightarrow \{0, 1\}$. We identify a truth assignment with the set $X' \subseteq X$ of variables that are assigned to 1 (true). Firstly, we define a necessary condition on the expressiveness of argumentation semantics which forces the problem to become complete.

*That is, negation symbols do only occur at variables. Note that the positive normal form can be established in polynomial time.*
Definition 27 (Reduction-suitable semantics, \text{sem-witness}) Let $X_1$ and $X_2$ be two non-empty and disjoint sets of the same size and $f : X_1 \to X_2$ a bijective mapping between these sets and let $x$ be an element not in $X_1 \cup X_2$. Moreover, let $X_1 \cup X_2 \cup \{x\} \subseteq Y$ for some set $Y$.

We call a semantics $\text{sem}$ over $Y$ reduction-suitable if for any sets $X_1, X_2, \{x\}$ satisfying the properties given above there is a coalitional framework whose size is polynomial in $Y$ such that $E \in \text{sem}(C\mathcal{F})$ iff (1) $E = X \cup \{x\}$, (2) $X \subseteq X_1 \cup X_2$ and (3) $\forall x \in X_1 \cup X_2$ ($x \in X$ iff $f(x) \notin X$).

Moreover, we call a coalitional framework that witnesses that a semantics is reduction-suitable a $\text{sem}$-witness coalitional framework.

A reduction-suitable semantics allows to describe all truth assignments of a formula in a compact way. The intuition is that $X_1$ and $X_2$ represent the variables and their negations, respectively. The set $x$ with $X \cup \{x\} \in \text{sem}(C\mathcal{F})$ encodes a truth assignment; i.e. the literals assigned 1. The variable $x$ is a technicality needed in the reduction. In the following proposition we make the observation that reduction-suitable semantics allow to represent exactly all truth assignments.

Proposition 16 Let $\varphi = \varphi(X)$ be a Boolean formula and let $X := \{\bar{x} \mid x \in X\}$ and let $\text{sem}$ be a reduction-suitable semantics over $Y$ where $X \cup X \cup \{x\} \subseteq Y$ and $x \notin X$. Then, there is a coalitional framework $C\mathcal{F}$ such that for each $T \in \text{sem}(C\mathcal{F})$, $T \cap X$ is a truth assignment of $\varphi$ and for each truth assignment $T$ of $\varphi$ there is a $Z \in \text{sem}(C\mathcal{F})$ such that $Z \cap X = T$.

Proof. Let $C\mathcal{F}$ be a $\text{sem}$-witness. Firstly, let $T \in \text{sem}(C\mathcal{F})$, clearly $T \cap X$ does only contain elements of $X$ and thus is a truth assignment. Now let, $T$ be a truth assignment of $\varphi$. We show that $Z \cup \{x\} \in \text{sem}(C\mathcal{F})$ with $Z := T \cup X'$, $X' := \{\bar{x} \mid x \notin T\}$ and $(Z \cup \{x\}) \cap X = T$. Clearly, we have that $Z \subseteq X \cup X$ and also that $(x \in Z$ if $\bar{x} \notin Z$ for all $x \in X \cup X$. This shows that both conditions of reduction-suitable semantics are satisfied.

In the following we introduce some notation to refer to subformulae of a formula and to the outermost logical Boolean connector $\lor$ or $\land$.

Definition 28 (Notation for subformulae of $\varphi$, relevance) Let $\varphi$ be a Boolean formula in positive normal form. We define $\text{ls}(\varphi)$ as the outermost logical connector in $\varphi$; that is,

$$\text{ls}(\varphi) := \begin{cases} \land & \text{if } \varphi = \psi_1 \land \psi_2 \\ \lor & \text{if } \varphi = \psi_1 \lor \psi_2 \\ \epsilon & \text{if } \varphi \text{ is a literal.} \end{cases}$$

Moreover, we define $\text{ls}(\varphi)$ (resp. $\text{rs}(\varphi)$) as the subformula on the left hand side (resp. right hand side) of $\text{ls}(\varphi)$ provided that $\text{ls}(\varphi) \neq \epsilon$. In the case of $\text{ls}(\varphi) = \epsilon$ we set $\text{ls}(\varphi) = \text{rs}(\varphi) = \varphi$. Note that we have that $\varphi = \text{ls}(\varphi)\text{ls}(\varphi)\text{rs}(\varphi)$ whenever $\text{ls}(\varphi) \neq \epsilon$.

Now, we can assign a string over $\{1, 2\}$ to refer to a subformula of $\varphi$, where 1 (resp. 2) stands for the left (resp. right) subformula wrt to the outermost logical connector. Formally, we define a function $\chi^\varphi$ from $\{1, 2\}^+ \cup \{0\}$ into the subformulae of $\varphi$ as follows (we write $\chi^\varphi_w$ for $\chi^\varphi(w)$ where $w \in \{1, 2\}^+ \cup \{0\}$):

$$\chi^\varphi_w := \begin{cases} \varphi & \text{if } w = 0 \\ \text{ls}(\varphi) & \text{if } w = 1 \\ \text{rs}(\varphi) & \text{if } w = 2 \\ \text{ls}(\chi^\varphi_w) & \text{if } w = x_1, x \in \{1, 2\}^+ \\ \text{rs}(\chi^\varphi_w) & \text{if } w = x_2, x \in \{1, 2\}^+. \end{cases}$$

Finally, we call a string $w \in \{1, 2\}^+$ relevant for $\varphi$ iff $\text{ls}(\chi^\varphi_w) \neq \epsilon$ for $x \in \{1, 2\}^*$ and $w = xi$ with $i \in \{1, 2\}$; or if $w = 0$. We will also just write $\chi_w$ if the formula $\varphi$ is clear from context.

Given the formula $\varphi = ((x_1 \land x_2) \lor \neg x_3) \land (\neg x_1 \lor x_3)$, for instance, we have that $\chi^\varphi_{11} = \neg x_1 \lor x_3$, $\chi^\varphi_{12} = x_2$, and $\chi^\varphi_{21} = \neg x_1$. 
We proceed with our reduction. Inspired by [15, 28] we construct a CGM corresponding to $\varphi(X)$ which essentially corresponds to the parse tree of $\varphi(X)$ and implements the game semantics of Boolean formulae (cf. [25]).

That is, we construct the CGM $\mathcal{M}(\varphi)$ corresponding to $\varphi(X)$ with $2 + 2|X|$ players: verifier $v$, refuter $r$, and agents $a_i$ and $\bar{a}_i$ for each variable $x_i \in X$. The CGM is turn-based, that is, every state is “governed” by a single player who determines the next transition. Each subformula $\chi_{i_1 \ldots i_k}$ of $\varphi$ has a corresponding state $q_{i_1 \ldots i_k}$ in $\mathcal{M}(\varphi)$ for $i_k \in \{0, 1\}$. If the outermost logical connective of $\varphi$ is $\land$, i.e. $let(\varphi) = \land$, the refuter decides at $q_0$ which subformula $\chi_i$ of $\varphi$ is to be satisfied (i.e. whether $\chi_1$ or $\chi_2$), by proceeding to the “subformula” state $q_i$ corresponding to $\chi_i$. If the outermost connective is $\lor$, the verifier decides which subformula $\chi_i$ of $\varphi$ will be attempted at $q_0$. This procedure is repeated until all subformulae are single literals. In the following we refer to the states corresponding to literals as literal states.

The difference from the construction from [28] is that formulae are in positive normal form (rather than CNF) and from [28, 15] in the way in that literal states are treated: literal states are governed by the agents $a_i$ or $\bar{a}_i$. The values of the underlying propositional variables $x, y$ are declared at the literal states, and the outcome is computed. That is, if $a_j$ executes $\top$ for a positive literal, i.e. $\chi_{i_1 \ldots i_k} = x_j$, at $q_{i_1 \ldots i_k}$, then the system proceeds to the “winning” state $q_\top$; otherwise, the system goes to the “sink” state $q_\bot$. Analogously, if $\bar{a}_j$ executes $\bot$ for a negative literal, i.e. $\chi_{i_1 \ldots i_k} = \neg x_j$, at $q_{i_1 \ldots i_k}$, then the system proceeds to the “winning” state $q_\top$; otherwise, the system goes to the “sink” state $q_\bot$.

Finally, the idea is to use $\text{sem}$-witness coalitional frameworks (for some reduction-suitable semantics) to represent all valuations of $\varphi(X)$ such that there is a “successful coalition” among these coalitions (representing the valuations of $\varphi$) iff $\varphi(X)$ is satisfiable. An example of the construction is shown in Figure 5.

Formally, the model is defined as follows.

**Definition 29 ($\mathcal{M}(\varphi)$)** Let $\varphi(X)$ be given. The model $\mathcal{M}(\varphi) = (\text{Agt}, Q, \Pi, \pi, \text{Act}, d, o, \zeta, \text{sem})$ is defined as follows:

1. $\text{Agt} := \{v, r\} \cup \{a_i, \bar{a}_i \mid x_i \in X\}$
2. $Q := \{q_0, q_\bot, q_\top\} \cup \{q_w \mid w \in \{1, 2\}^+\}$ relevant for $\varphi$
3. $\Pi := \{\text{sat}\}$
Inteligencia Artificial 46(2010) 63

- \( \pi(q_T) = \{ \text{sat} \} \)
- \( \text{Act} := \{ 1, 2, \top, \bot \} \)
- \( d_v(q_w) = \{ 1, 2 \} \) for each \( w \) relevant for \( \varphi \) with \( \text{lcl}(\chi_w^{v}) = \land \); \( d_v(q_w) = \{ 1, 2 \} \) for each \( w \) relevant for \( \varphi \) with \( \text{lcl}(\chi_w^{v}) = \lor \); \( d_w(q_w) = \{ \top, \bot \} \) for each \( w \) relevant for \( \varphi \) with \( \chi_w = x_i \); \( d_{\text{a}}(q_w) = \{ \top, \bot \} \) for each \( w \) relevant for \( \varphi \) with \( \chi_w = \neg x_i \); \( d_{\text{a}}(q) = \{ \top \} \) for \( q \in \{ q_T, q_\bot \} \) and \( x \in \text{agt} \).

- Note that the model is turn-based (except for the states \( q_T \) and \( q_\bot \)); that is, in each state only one agent can execute actions. Hence, transitions do only depend on single actions and not on action profiles: \( o(q_0)(1) = q_1, o(q_0)(2) = q_2, o(q_w)(1) = q_{w1} \) if \( w \) is relevant for \( \varphi \); \( o(q_w)(2) = q_{w2} \) if \( w \) is relevant for \( \varphi \); \( o(q_w)(\top) = q_T \) (resp. \( o(q_w)(\bot) = q_\bot \) if \( \chi_w \in X \); \( o(q_w)(\top) = q_\bot \) (resp. \( o(q_w)(\bot) = q_T \)) if \( \neg \chi_w \in X \).

- \( \text{scm} \) is a reduction-suitable semantics over \( \text{agt} \). Finally, we set \( \zeta(A)(q) = \text{CF} \) where \( \text{CF} \) is some \( \text{scm} \)-witness coalitional framework for the sets \( X_1 := \{ a_i | x_i \in X \} \), \( X_2 := \{ \bar{a}_i | x_i \in X \} \), and \( x := \{ v \} \) for all \( A \subseteq \text{agt} \) and \( q \in Q \).

We define a \( v \)-choice of \( \mathcal{M}(\varphi) \) as “the graph” that occurs if from states controlled by the verifier \( v \) all transitions but one are removed (i.e. at each state controlled by the verifier it does only have one action to execute). With other words, we fix a strategy of \( v \). The following lemma is essential for our reduction.

**Lemma 17** Let \( \varphi(X) \) be a Boolean formula in positive normal form.

(a) If \( T \) is a satisfying truth assignment of \( \varphi \), then there is a \( v \)-choice of \( \mathcal{M}(\varphi) \) such that for the set \( L \) of all literal states reachable from \( q_0 \) it holds that \( \{ x \in X \mid q_w \in L \text{ and } \chi_w = x \} \subseteq T \) and \( \{ x \in X \mid q_w \in L \text{ and } \chi_w = x \} \cap T = \emptyset \).

(b) If there is a \( v \)-choice of \( \mathcal{M}(\varphi) \) such that for the set \( L \) of all literal states reachable from \( q_0 \) we have that for any \( q_w, q_\ell \in L \) the formula \( \chi_w \land \chi_\ell \) is satisfiable (i.e. there are no complementary literals) then the set \( \{ x \in X \mid q_w \in L \text{ and } \chi_w = x \} \) is a satisfying truth assignment of \( \varphi \).

**Proof.**

(a) Let \( T \) be a satisfying truth assignment of \( \varphi \). We construct a \( v \)-choice with the stated property. Firstly, we call a state \( q_w \) a \( v \)-state (resp. \( r \)-state) if \( q_w \) is controlled by \( v \) (resp. \( r \)). Next, we consider a state \( q_w \) such that \( q_{w1} \) and \( q_{w2} \) are literal states. Firstly, let us consider the case that \( q_w \) is a \( v \)-state. Then, we label it with \( \top \) if for some \( i \) we have that \( x_i \in T \) (resp. \( x_i \notin T \)) if \( \chi_{wi} = x \) (resp. \( \chi_{wi} = \neg x \)) for \( x_i \in X \); otherwise we label the state with \( \bot \). Secondly, if \( q_w \) is a \( r \)-state we label it with \( \top \) if the above condition holds for both \( i \); otherwise we label it with \( \bot \).

In the second step, we label each \( v \)-state with two labelled successor states with \( \top \) if at least one of its successors has the label \( \top \); otherwise we assign to it the label \( \bot \). On the other hand, if it is an \( r \)-state, if both successors have the label \( \top \) we label it \( \top \); otherwise we assign to it the label \( \bot \).

We proceed like this until all states \( q_w \) with a relevant \( w \) are labelled.

Now, in lexicographical order we go through all \( w \) relevant for \( \varphi \) and perform the following steps. If \( q_w \) is a \( v \)-state reachable from \( q_0 \), we remove the transition to the successor which is labelled \( \top \) and if both transitions are labelled \( \top \) we remove any of its transitions. Else, if \( q_w \) is a \( v \)-state and not reachable from \( q_0 \) we remove any of its transitions. Following this procedure, we have to show that there is no reachable \( v \)-state such that both successors are labelled \( \bot \). Suppose such a state \( q_w \) is reachable. Since the verifier has no strategy to avoid this state (note that \( w \) is lexicographically minimal) the disjunction \( \chi_{w1} \lor \chi_{w2} \) must be true under \( T \) in order to make \( \varphi \) true. However, since both states \( q_{w1} \) and \( q_{w2} \) are labelled \( \bot \) the truth assignment falsifies \( \chi_{w1} \) and \( \chi_{w2} \) (that can easily be seen by induction). Contradiction!

Hence, the construction yields a \( v \)-choice. We need to show that there is a \( v \)-choice such that \( S := \{ x \in X \mid q_w \in L \text{ and } \chi_w = x \} \) is a subset of \( T \). Firstly, assume there is some \( x \in S \) not in \( T \). Let \( q_{wi} \) be some reachable literal state with \( \chi_{wi} = x \). By construction of the \( v \)-choice, this
implies that the state $q_w$ is an $r$-state. (Otherwise, the other alternative would have been selected (according to the labelling algorithm); or, both successors of $q_w$ would be labelled $\bot$ what is not possible as shown above.) But this means, that the formula $\chi_{w1} \land \chi_{w2}$ needs to be satisfied and hence $x \in T$. Secondly, assume that $x \in \{x \in X \mid q_w \in L$ and $\chi_w = \neg \alpha\} \cap T$ and let $q_w$ be the state reachable with $\chi_w = \neg \alpha$. Following the same reasoning as above the state $q_w$ must be an $r$-state and this contradicts the fact the the formula $\chi_{w1} \land \chi_{w2}$ must be true under $T$.

(b) Let us consider a $v$-choice of $M(\varphi)$ with the stated property and suppose that $T := \{x \in X \mid q_w \in L$ and $\chi_w = x\}$ is not a satisfying truth assignment of $\varphi$. Note that the $v$-choice corresponds to the selection of the left (resp. right) hand subformula of any subformula $\chi_w$ with $lc(\chi_w) = v$. We say that a subformula $\chi_w$ is reachable if the state $q_w$ is reachable in the very $v$-choice. Since $T$ is not a satisfying truth assignment there is some reachable subformula (possibly $\varphi$ itself) not satisfied by $T$. Let $\chi_w$ be such a reachable subformula with lexicographically maximal $w$ (so, it is a relative “small” subformula). We consider two cases. Firstly, suppose that $\chi_w$ is not satisfiable. Then, due to the maximality of $w$, $\chi_w$ must be a conjunction of literals among that are two complementary ones, we denote them by $\chi_i$ and $\chi_{i'}$ (otherwise it would be satisfiable); hence, we have that $q_v, q_{i'} \in L$. Contradiction!

So, suppose $\chi_w$ is satisfiable but false under $T$. Then, due to the maximality of $w$, $\chi_w$ can either be $x_i$ for some $x_i \notin T$ or be $\neg x_j$ for some $x_j \in T$. Suppose $\chi_w = x_i$; then $q_w$ is reachable and thus $x_i \in T$. Contradiction! On the other hand, if $\chi_w = \neg x_j$ then $q_w \in L$ and since $x_j \in T$ there is some other state $q_v \in L$ with $\chi_v = x_j$ and $\chi_w \land \chi_v$ not satisfiable. Contradiction!

Similar to [15, 28], we have the following result which shows that the construction above is a polynomial-time reduction of $\text{SAT}$ to model checking $\text{COALATL}$.

**Proposition 18** The model $M(\varphi)$ is constructible in polynomial-time wrt the size of $\varphi$ and we have that

\[ \varphi(X) \text{ is satisfiable if, and only if, } M(\varphi), q_0 \models \langle \forall \rangle \diamond \text{sat}. \]

**Proof.** Firstly, we analyze the construction to show that its length is polynomially wrt the input. The number of agents is polynomially in $X$; the number of states is polynomially in the number of subformulae of $\varphi$ and for each state there are at most two transitions (apart form the states $q_1$ and $q_T$).

$\Rightarrow$ Let $\varphi(X)$ be satisfiable and let $T \subseteq X$ be a satisfying truth assignment. By construction of $M(\varphi)$ we have that $C := \{v, a_i, a_j \mid x_i \in T, x_j \notin T \} \in \text{sem}(\{\langle v \rangle\}(q_0))$; hence, it suffices to show that $M(\varphi), q_0 \models \langle C \rangle \diamond \text{sat}$. Now by Lemma 17(a) there is a strategy $s_v$ of $v$ such that for the set $L$ of all literal states reachable from $q_0$ we have that $S := \{x \mid q_w \in L$ and $\chi_w = x\} \subseteq T$ and $\{x \mid q_w \in L$ and $\chi_w = \neg x\} \cap T = \emptyset$. Hence, for any $x_i \in S$ the agent $a_i$ is in $C$ and for any $x_i \in \{x \mid q_w \in L$ and $\chi_w = \neg x\}$ the agent $a_i$ is in $C$. Finally, the strategy $s_v$ of $a \in C \setminus \{v\}$ is to execute $T$ if $a = a_i$ and $\bot$ if $a = a_i$ in the states controlled by $a$. The complete strategy profile consisting of $s_b$ for $b \in C$ ensures $\diamond \text{sat}$.

$\Leftarrow$: Suppose that $M(\varphi), q_0 \models \langle v \rangle \diamond \text{sat}$. Then there is a coalition $C \in \text{sem}(\{\langle v \rangle\}(q_0))$ such that $M(\varphi), q_0 \models \langle C \rangle \diamond \text{sat}$. By Lemma 17(b) it remains to show that for the set $L$ of all literal states reachable from $q_0$ we have that for any $q_w, q_v \in L$ the formula $\chi_w \land \chi_v$ is satisfiable. Suppose the contrary; that is, there are $q_w, q_v \in L$ such that $\chi_w \land \chi_v \equiv \bot$; say $\chi_w = x_i$. By Proposition 16 $C \setminus \{v\}$ does correspond to a truth assignment of $\varphi$; hence, either $a_i \in C$ or $\bar{a}_i \in C$. Hence, if $a_i \in C$ then $a_i \notin C$ and the opponents (to whom $a_i$ belongs) has a strategy to reach $q_{\bot}$ from $q_w$. But this contradicts $M(\varphi), q_0 \models \langle C \rangle \diamond \text{sat}$. The same reasoning is applied if $a_i \in C$ and $\bar{a}_i \notin C$.

The following result is obvious, we can reduce $\text{SAT}$ and $\text{UNSAT}$ to model checking $M(\varphi), q_0 = \{\text{sat}\}$ and $M(\varphi), q_0 = \{\neg \text{sat}\}$, respectively.

**Theorem 19** Model checking $\text{COALATL}$ is $\text{NP}$-hard and $\text{coNP}$-hard for any reduction-suitable semantics.

### 6.3 $\Delta_2^P$-hardness

Finally, we show our main result, the $\Delta_2^P$-hardness of model checking $\text{COALATL}$ for reduction-suitable semantics. We do so by reducing $\text{SN SAT}$, a typical $\Delta_2^P$-complete problem (cf. [32]).
**Definition 30 (snsat)**

**Input:** \( p \) sets of propositional variables \( X^r = \{x^r_1, \ldots, x^r_k\} \); \( p \) propositional variables \( z^r \) and \( p \) Boolean formulae \( \varphi^r \) in positive normal form (i.e., negation is allowed only on the level of literals) for \( r = 1, \ldots, p \). Each \( \varphi^r \) involves only variables from \( X^r \cup \{z^1, \ldots, z^{r-1}\} \), with the following requirement: \( z^r = \exists Y^r(\varphi^r(z^1, \ldots, z^{r-1}, Y^r)) \) where \( Y^r \subseteq X^r \). (The notation is understood as follows: there is a truth assignment assigning 1 (resp. 0) to the variables in \( Y^r \) (resp. \( X^r \setminus Y^r \)) such that \( \varphi^r(z^1, \ldots, z^{r-1}, X^r) \) is true under this assignment.)

**Output:** The value of \( z^p \).

We use \( I = (\varphi_1(X^1), \ldots, \varphi_p(X^p)) \) or just \( I = (\varphi_1, \ldots, \varphi_p) \) to denote an instance of snsat; and set \( Z = \{z^1, \ldots, z^p\} \).

We often need to analyze a solution of an snsat-instance. In the following we show how a solution can formally be stated.

**Definition 31 (Witness and solution of an snsat-instance)**

Let \( I = (\varphi_1, \ldots, \varphi_p) \) be an snsat-instance. A tuple \((T_1, \ldots, T_p)\) is an I-witness if it satisfies the following properties:

1. \( T_i \subseteq \{z^1, \ldots, z^i\} \cup X^i \);
2. If \( \varphi_i \) is satisfiable under the partial assignment \( \{z^{j} \in T_{j} \mid j < i\} \) then \( T_i \) is a satisfying truth assignment of \( \varphi_i \) and \( z^i \in T_i \); else \( z^{i} \not\in T_{i} \) and \( T_i \cap X^i = \emptyset \);
3. \( z^{i} \in T_i \Rightarrow z^{i} \in T_j \) for all \( j \geq i \),

An I-witness \( T \) is a solution of \( I \) iff \( z^p \in T_p \).

In the next proposition we state some properties about solutions and witnesses.

**Proposition 20**

Let \( I = (\varphi_1, \ldots, \varphi_p) \) be an snsat-instance and let \( T = (T_1, \ldots, T_p) \) be an I-witness.

(a) Let \( T' = (T'_1, \ldots, T'_p) \) be another I-witness; then, \( T'_i \cap Z = T_i \cap Z \) for all \( i = 1, \ldots, p \). (That is, among all witnesses the values for the \( z^i \)'s are uniquely determined.)

(b) If \( I' = (\varphi_1, \ldots, \varphi_i) \) with \( i < p \) has a solution, then \( T = (T_1, \ldots, T_i) \) is a solution of \( I' \).

(c) Let \( I' = (\varphi_1, \ldots, \varphi_p, \varphi_{p+1}) \) be such that \( I' \) has no solution; then, \( T' = (T_1, \ldots, T_p, \{z \in T_p\}) \) is an \( I' \)-witness.

**Proof.**

(a) Suppose that \( i \) is the minimal index for which both sets \( T_i \cap Z \) and \( T'_i \cap Z \) differ; that is, wlog, \( z_i \in T_i \) and \( z_i \not\in T'_i \). This is a contradiction to property (2) of the definition of a witness; as the satisfiability of \( \varphi_i \) under \( z_1, \ldots, z_{i-1} \) is uniquely determined.

(b) Follows from (a).

(c) We show that \( R := \{z \in T_p\} \) satisfies the conditions (1-3) in the Definition of a witness. Clearly, \( R \subseteq \{z^1, \ldots, z^{p+1}\} \). Since \( I' \) has no solution, there is no way to make \( z_{p+1} \) true. Clearly, \( z_{p+1} \not\in R \) and \( R \cap X^{p+1} = \emptyset \). Finally, condition (3) is satisfied by definition of \( R \).

Our reduction of snsat_1 is a modification of the reduction of snsat_2 presented in [15, 28] and extends the NP/coNP-hardness construction of the previous section. Consider an snsat_1 instance \( I = (\varphi_1, \ldots, \varphi_p) \). Essentially, we construct models \( M(\varphi_r) \) for \( r = 1, \ldots, p \) as shown above but we label each state of \( M(\varphi_r) \) and each agent name by an additional superindex \( r \) (that is, states are denoted by \( q_{i_1}^r \) and agents by \( a_1^r \) and \( a_2^r \)). The main difference is how the literal states corresponding to literals \( z_i \) and \( \neg z_i \) are treated. We connect such states of model \( M(\varphi_r) \) with \( r > 1 \) with the initial state \( q_0^{-1} \) of model \( M(\varphi_r-1) \). Additionally, states referring to negated variables \( z_i \) are labeled with a proposition \( \text{neg} \). Finally, the full model \( M(I) \) wrt an snsat_1 instance \( I \) is given by the combination of these models as just explained. In particular, the set of agents is given by \( \{v, r\} \cup \{a_1^r, a_2^r \mid x^r_i \in \bigcup_{r=1}^p X^i\} \). An example of
the construction is shown in Figure [6]. Given the model $M(I)$ we use $M(\phi_i)$ to refer the restriction of $M(I)$ to the states with superindex $i$. In Figure [6] the two submodels $M(\phi_1)$ and $M(\phi_2)$ are framed.

The formulae used in this reduction are more sophisticated as they have to account for the nested structure of an $\text{SNSAT}_1$ instance. We define

$$\phi_0 \equiv \top$$

and

$$\phi_r \equiv \langle v \rangle (\lnot \text{neg } \cup (\text{neg } \land \langle \emptyset \rangle \circ \lnot \phi_{r-1}))$$

for $r = 1, \ldots, p$.

Before we come to the theorem proving the reduction we state a fundamental lemma which can be seen as counterpart of Lemma [17].

**Lemma 21** Let $I = (\varphi_1, \ldots, \varphi_p)$ be an $\text{SNSAT}_1$ instance.

(a) Let $T = (T_1, \ldots, T_p)$ be a solution for $I$. For all $r = 1, \ldots, p$, if $z_r \in T_r$ then there is a $v$-choice of $M(\varphi_r)$ such that for the set $L$ of all literal states reachable from $q_0^r$ and which belong to $M(\varphi_r)$ it holds that $\{x \in X^r \cup Z \mid q_w \in L \text{ and } \chi_w^{q_w} = x\} \subseteq T_r$ and $\{x \in X^r \cup Z \mid q_w \in L \text{ and } \chi_w^{q_w} = \lnot x\} \cap T_r = \emptyset$.

(b) Let $T^{-1} = (\varphi_1, \ldots, \varphi_{p-1})$ and let $T^{p-1} = (T_1, \ldots, T_{p-1})$ be an $T^{p-1}$-witness. Then, if there is a $v$-choice of $M(\varphi_p)$ such that for the set $L$ of all literal states reachable from $q_0^p$ that belong to $M(\varphi_p)$ we have that

\[
M(I) \quad M(\varphi_2(z_1, X^2)) \quad M(\varphi_1(X^1))
\]
(i) for any \( q_w, q_v \in L \) the literals \( \chi^q_w \) and \( \chi^q_v \) are non-complementary; and
(ii) if \( q_v \in L \) with \( \chi^q_v = z_i \) (resp. \( \chi^q_v = \neg z_i \)) then \( z_i \in T_i \) (resp. \( z_i \notin T_i \));
then, \( T = (T_i^{r-1}, \ldots, T_0^{p-1}, T_p) \) is a solution for \( I \) where
\[
T_p = \{ x \in X^r \mid q_w \in L \text{ and } \chi_w = x \} \cup \{ z^i \mid z^i \in T_i^{p-1}, i < p \} \cup \{ z^p \}.
\]
(c) \( M(I), q_0^1 \models \phi_i \) if, and only if, \( M(I), q_0^1 \models \phi_j \); and \( M(I), q_0^1 \models \neg \phi_i \) if, and only if, \( M(I), q_0^1 \models \neg \phi_j \)
for all \( j \geq i \).

Proof.

(a) Let \( z_r \in T_r \). As in the proof of Lemma 17(a) we proceed with the following labeling: We consider a state \( q_w \) such that \( q_w \) is a \( \nu \)-state. Then, we label it with \( \top \) if for some \( i \) with \( \chi_{w_i} = x \in X^r \cup Z \) (resp. \( \chi_{w_i} = \neg x \), \( x \in X^r \cup Z \)) we have that \( x \in T_r \) (resp. \( x \notin T_r \)); otherwise we label the state with \( \bot \). Secondly, if \( q_w \) is a \( \tau \)-state we label it with \( \top \) if the above condition holds for both \( i \); otherwise by \( \bot \). In the second step, (labeling non-literal states) we proceed exactly as before.
In the second step, we label each \( \nu \)-state with two labelled successor states with \( \top \) if at least one of its successors has the label \( \top \); otherwise we assign to it the label \( \bot \). On the other hand, if it is a \( \tau \)-state, if both successors have the label \( \top \) we label it \( \top \); otherwise we assign to it the label \( \bot \). We proceed like that until all states \( q_w \) with a relevant \( w \) are labelled.
The \( \nu \)-choice is constructed in the very same way as in Lemma 17(a) and also the verification that the \( \nu \)-choice has the same properties is done in the very same way.

(b) Let us consider a \( \nu \)-choice with the stated properties (i) and (ii). We adopt the notation from the proof of Lemma 17(b). By definition of \( T_p \) it is obvious that it satisfies conditions 1. and 3. of Definition 31.
Suppose that \( T \) is not a solution of \( I \); that is, condition 2. of Definition 31 is violated. Since \( T^{p-1} \) is a witness of \( I^{p-1} \) there is a reachable subformula \( \chi_w \) of \( \varphi_p \) with lexicographically minimal \( w \) that is not satisfiable given the valuation of the \( z \)'s according to \( T^{p-1} \) (note that the \( z \)'s are uniquely determined in \( T^{p-1} \)). Note, that the condition (ii) guarantees that the “right” value is chosen for \( z_i \in T_i \) and \( z_i \notin T_i \). The rest of the proof is done analogously to the one given in Lemma 17(b) by considering the literal states corresponding to variables \( X^p \cup Z \). This proves that \( T \) is a solution.

c) Along each path from \( q_0^1 \) to a state labelled sat there are at most \( i - 1 \) states labelled neg. Hence, the truth of a formula \( \phi_j \) with \( j \geq i \) is equivalent to the truth of \( \phi_i \).

Similar to [15, 28], we have the following result which shows that the construction above is a polynomial-time reduction of \( \text{snsat}_1 \) to model checking \( \text{coalatl} \).

**Theorem 22** The size of \( M(I) \) and of the formulae \( \phi_p \) is polynomially in the size of the \( \text{snsat}_1 \) instance \( I = (\varphi_1, \ldots, \varphi_p) \) and we have the following:

There is a solution \( T = (T_1, \ldots, T_r) \) of \( I^r = (\varphi_1, \ldots, \varphi_r) \) if, and only if, \( M(I), q_0^1 \models \phi_i \)
for \( l \geq r \) and \( r \leq p \).

Proof. That \( M(I) \) is polynomially wrt \( \varphi \) follows from Proposition 18 and by the way it is constructed. The size of \( \phi_p \) also is polynomially in \( \varphi(X) \).

The proof is done by induction on \( r \).

**Induction start.** The induction starts with \( r = 1 \). An \( \text{snsat}_1 \)-instance with \( r = 1 \) is given by \( \exists_1 \varphi(X^1) \). This corresponds to the satisfiability problem. So, this case is proven in Proposition 18 and Lemma 21(c).
Induction step. Suppose the claim holds up to $r < p$. We show that there is a solution $T = (T_1, \ldots, T_{r+1})$ of $I^{r+1}$ if $\mathcal{M}(I^{r+1})_0 \models \phi_{r+1}$.

$\Rightarrow$: Let $T = (T_1, \ldots, T_{r+1})$ be a solution for $I^{r+1}$.

From the reduction-suitableness we have that there is a coalition $C \in \text{sem}(\zeta(\{v\})(q_0^r))$ with

$$C = \{v\} \cup \bigcup_{i=1}^r \{a_i^j : a_i^j \in T_i \} \cup \bigcup_{i=1}^{r+1} \{a_i^j : x_i^j \not\in T_i \}$$

Note that, if $z_i \in T_i$ then $T_i = \{x_j^i \mid a_i^j \in C, j = 1, \ldots, s \} \cup \{z_i \mid \mathcal{M}(I^{r+1}), q_0^r \models \phi_j, j = 1, \ldots, i \}$.

We show that $\mathcal{M}(I^{r+1})_0 \models \langle C \rangle (\neg \text{negsat} (\text{sat} \lor (\text{neg} \land \langle \emptyset \rangle \lor \neg \phi_{r+1})))$. Let $s_i^{r+1}$ be the partial strategy corresponding to the v-choice according to Lemma 21(a). The set of reachable literal states $L^{r+1}$ in $\mathcal{M}(\varphi_{r+1})$ corresponds to finite sequences of states starting in $q_0^{r+1}$. In the following, we call such literal states referring to some $z^i$ (resp. $\neg z^i$), $z$ (resp. $\neg z$)-literal states and the others $x$ (resp. $\neg x$)-literal states. Let $\lambda$ be a finite sequence of states starting in $q_0^{r+1}$ and ending in one of the states in $L^{r+1}$. Then this sequence has one of the following properties: (1) The last state is an $x$, $z$, $\neg x$-literal state and there is no $\neg z$-literal state on it; or (2) the last state is a $\neg z$-literal state and there is no other $\neg z$-literal state on it. Note, that due to Lemma 21(a) it is not possible to reach two complementary literal states. We show, that we can extend $s_v$ to a strategy $s_C$ that witnesses the truth of $\phi_{r+1}$.

Case 1. For a path ending with an $x$-literal state $\chi_w = x_i^{r+1}$ we have that $x_i^{r+1} \in T_{r+1}$ and thus $a_i^{r+1} \in C$. If this agent executes $\top$ the next state is $q_{r+1}^\top$.

Case 2. For a path ending with an $\neg x$-literal state $\chi_w = \neg x_i^{r+1}$ we have that $x_i^{r+1} \not\in T_{r+1}$ and thus $\neg a_i^{r+1} \in C$. If this agent executes $\bot$ the next state is $q_{r+1}^\bot$.

Case 3. For a path ending with an $z$-literal state $\chi_w = z_i^{r+1}$ we have that $z_i^{r+1} \in T_{r+1}$ hence also $z_i^{r+1} \in T_i$. By induction hypothesis we have that $\mathcal{M}(I), q_0^r \models \phi_r$. Now, it is easily seen by another induction that the very same coalition $C$ has a winning strategy $s_C'$ witnessing $\phi_r$. We combine the strategy $s_v$ with $s_C'$.

Case 4. For a path ending with an $\neg z$-literal state $\chi_w = \neg z_i^{r+1}$, we have that $z_i^{r+1} \not\in T_{r+1}$ hence also $z_i^{r+1} \not\in T_i$. By induction hypothesis we have that $\mathcal{M}(I), q_0^r \not\models \phi_r$; hence, $\mathcal{M}(I^{r+1}), q_w \models \neg \text{neg}\land \langle \emptyset \rangle \lor \neg \phi_{r+1}$.

These cases give us the desired strategy $s_C$ witnessing $\phi_{r+1}$ for the chosen coalition $C$.

$\Leftarrow$: Suppose that $\mathcal{M}(I^{r+1}), q_0^{r+1} \models \phi_{r+1}$.

Let $z^i$ with $i < r + 1$ and with an maximal index $i$ be such that $\mathcal{M}(I^{r+1}), q_0^{r+1} \models \phi_i$. Then, there is a solution $T^* = (T_1, \ldots, T_i) \top$ of $I^r$ by induction hypothesis. According to Proposition 20(b) and (c) we can extend this solution to an $I^r$ witness $T^* = (T_1, \ldots, T_r)$ of $I^r$.

Now, since $\mathcal{M}(I^{r+1}), q_0^{r+1} \models \phi_{r+1}$, there is a coalition $C \in \text{sem}(\zeta(\{v\})(q_0^{r+1}))$ and a strategy $s_C$ that witnesses the truth of the formula. Let $L$ be the set of all reachable literal states in $\mathcal{M}(\varphi_{r+1})$ under $s_C$ (a v-choice is contained implicitly). We show that the preconditions (i) and (ii) of Lemma 21(b) are satisfied by this v-choice.

Condition (i). Firstly, for each reachable $x/\neg x$-literal state $q_w^{r+1} \in L$ with $\chi_w^{r+1} = x$ we must have that $a_i^{r+1}$ (resp. $\neg a_i^{r+1}$) is in $C$ if $x = x_i^{r+1}$ (resp. $x = \neg x_i^{r+1}$) (otherwise the formula $\neg \text{negsat}$ would not be true). Hence, there cannot be any complementary $x$-literal states contained in $L$ due to the reduction-suitable semantics (again, otherwise $\neg \text{negsat}$ would not be true).

Secondly, observe that for any positive $z$-literal state $q_w^{r+1} \in L$ with $\chi_w^{r+1} = z^i$ we have that $\mathcal{M}(I^{r+1}), q_w^{r+1} \models \phi_{r+1}$ and by Lemma 21(c) also $\mathcal{M}(I^{r+1}), q_w^{r+1} \models \phi_i$; hence, by induction hypothesis, $z^i \in T_i$.

For each negative $z$-literal state $q_w^{r+1} \in L$ with $\chi_w^{r+1} = \neg z^i$ we have that $\mathcal{M}(I^{r+1}), q_0^{r+1} \models \neg \phi_i$ and thus, by Lemma 21(c) $\mathcal{M}(I^{r+1}), q_w^{r+1} \models \neg \phi_i$. By induction hypothesis, we have that for any $I^r$-witness $(T_1^r, \ldots, T_i^r)$ it holds that $z_i^r \not\in T_i^r$.
Condition (ii). Let $q_0\in L$ with $\chi^{q_0} = z_i$. We show that $z_i \in T_i$. Suppose the contrary. Then, $M(I^{r+1}), q_0 \models \neg \phi_i$ and by Lemma 21(c) also $M(I^{r+1}), q_0 \models \neg \phi_{r+1}$. However, since along the path from $q_0^{r+1}$ to $q_0'$ we always have $\neg \neg \neg \phi$ the coalition $C$ performing strategy $s_C$ witnesses the formula $\phi_{r+1}$ in $q_0'$. Contradiction!

Secondly, assume $q_v \in L$ with $\chi^{q_v} = \neg z_i$ and again, by the sake of contradiction, that $z_i \in T_i$. Then $M(I^{r+1}), q_0 \models \phi_i$ and thus also $M(I^{r+1}), q_0' \models \phi_i$ by Lemma 21(c). However, on the path to $q_v \in L$ we have visited a state labeled neg; hence, we must have $M(I^{r+1}), q_0 \models \neg \phi_i$. Contradiction!

Now by Lemma 21(b), we can construct a solution for $I^{r+1}$ form $T^r$.

The reduction gives us the following hardness result.

**Theorem 23** Model checking CoalATL is $\Delta_2^P$-hard for any reduction-suitable semantics.

With Corollary 11, we obtain the following completeness result.

**Theorem 24** Model checking CoalATL is $\Delta_2^P$-complete for any reduction-suitable semantics $\text{sem}$ with $\forall E \in \text{sem \in NP}$.

Finally, we show completeness for the stable semantics.

**Corollary 25** Model checking CoalATL is $\Delta_2^P$-complete for the $\text{sem}_{\text{stable}}$ semantics.

**Proof.** We have to show that the semantics is reduction-suitable; that is, we have to come up with a $\text{sem}$-witness over $\text{agt}$.

Consider the coalitional framework shown in Figure 7. We show that this is a witness for the reduction suitableness of the stable semantics. Let $X = \{x_1, \ldots, x_n\}$ and $\bar{X} = \{\bar{x}_1, \ldots, \bar{x}_n\}$ be given and $f(x_i) := \bar{x}_i$. We show that the conditions from Definition 27 are met.

"$\Rightarrow$": Let $E \in \text{sem}_{\text{stable}}$. Clearly, $r \notin E$ because $\{r\}$ is not conflict-free. Moreover, $\{\bar{v}\} \in E$; as $\{v\}$ cannot be attacked by $E$. We also have that $E \cap \bar{X} = \emptyset$. Suppose there is an $i$ such that $\{a_i, \bar{a}_i\} \cap E = \emptyset$. Then, $\{a_i, \bar{a}_i\}$ is not attacked by $E$. Contradiction. Now suppose that $\{a_i, \bar{a}_i\} \subseteq E$. This is a contradiction since $\{a_i, \bar{a}_i\}$ is not conflict-free.

"$\Leftarrow$": Suppose $E \cup \{v\}$ satisfies condition (1), (2), and (3). Clearly, $E \cup \{\bar{v}\}$ is conflict-free. Because for each $i$ either $a_i$ or $\bar{a}_i$ is in $E \cup \{v\}$ every element outside is attacked by some element from $E \cup \{v\}$.

7 Related and Future Work

**Related Work.** The main inspiration for our work was the powerful argument-based model for reasoning about coalition structures proposed by Amgoud [3]. Indeed, our notion of coalitional framework (Def. 1) is based on the notion of framework for generating coalition structures (FCS) presented in Amgoud’s paper. However, in contrast with Amgoud’s proposal, our work is concerned with extending ATL by argumentation in order to model coalition formation.
Types of Semantics | MC Complexity
---|---
\( \forall \mathcal{E} \mathcal{R}_{sem} \in C \) | \( P^{NP^c} \)
\( sem_{preferred} \) | \( \Delta^p_2 \)
reduction-suitable | \( \Delta^p_2 \)-hard
\( \forall \mathcal{E} \mathcal{R}_{sem} \in \text{NP}, \text{reduction-suitable} \) | \( \Delta^p_2 \)-complete
\( sem_{stable} \) | \( \Delta^p_2 \)-complete
\( \forall \mathcal{E} \mathcal{R}_{sem} \in \text{NP} \) | \( \Delta^p_2 \)
\( sem_{admissible}, sem_{complete} \) | \( \Delta^p_2 \)
polynomially many coalitions, polynomially enumerable \( sem \) | \( P \)-complete
\( sem_{grounded} \) | \( P \)-complete

Figure 8: Overview of the model checking results modulo the complexity of the used argumentation semantics.

Previous research by Hattori et al. [24] has also addressed the problem of argument-based coalition formation, but from a different perspective than ours. In [24] the authors propose an argumentation-based negotiation method for coalition formation which combines a logical framework and an argument evaluation mechanism. The resulting 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.

Modelling argument-based reasoning with bounded rationality has also been the focus of previous research. In [39] the authors propose the use of a framework for argument-based negotiation, which allows for a strategic and adaptive communication to achieve private goals within the limits of bounded rationality in open argumentation communities. In contrast with our approach, the focus here is not on extending a particular logic for reasoning about coalitions, as in our case. Recent research in formalizing coalition formation has been oriented towards adding more expressivity to Pauly’s coalition logic [34]. 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 \( \phi \).” In [9], 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.

It must be noted that the adequate formalization of preferences has deserved considerable attention within the argumentation community. In preference-based argumentation theory, an argument may be preferred to another one when, for example, it is more specific, its beliefs have a higher probability or certainty, or it promotes a higher value. Recent work by Kaci et al. [29, 30] has provided interesting contributions in this direction, including default reasoning abilities about the preferences over the arguments, as well as an algorithm to derive the most likely preference order.

Future Work. Indeed, in the line with the previous paragraph, one of our future research lines is to extend our current formalization of \( \text{COALATL} \) to capture more complex issues in preference handling and to consider more sophisticated semantics.

In the semantics presented in Definition 15 a valid coalition is initially formed and kept until the property is fulfilled. For instance, consider formula \( \langle A \rangle \square \varphi \). The formula is true in \( q \) if a valid coalition \( B \) in \( q \) can be formed such that it can ensure \( \square \varphi \). On might strengthen the scenario and require that \( B \) must be valid in each state on the path \( \lambda \) satisfying \( \varphi \). Formally, the semantics could be given as follows:

\( q \models \langle A \rangle \square \varphi \) if, and only if, \( q \models \varphi \) and there is a coalition \( B \in \text{vc}(q, A) \) and a common strategy \( s_B \in \Sigma_B \) such that for all paths \( \lambda \in \text{out}(q, s_B) \) and for all \( i \in \mathbb{N} \) it holds that \( \lambda[i] \models \varphi \) and \( B \in \text{vc}(\lambda[i], A) \). The
last part specifies that $B$ must be a valid coalition in each state $q_i = \lambda[i]$ of $\lambda$.

In the semantics just presented the formed coalition $B$ must persist over time until $\varphi$ 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. At first, $B$ must be a valid coalition in state $q$ leading to a state in which $\varphi$ 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.

Finally, we also aim at an actual implementation of a subset of $\text{CoalATL}$ in order to perform experiments to assess our proposal when modelling complex problems. We will consider a restriction to some particular argumentation semantics for which proof procedures can be deployed, such as assumption-based argumentation [23]. Research in this direction is currently being pursued. Also the model checking complexity of the goal-based semantics still has to be determined.

8 Conclusions

In this paper we have presented $\text{CoalATL}$, an extension of $\text{ATL}$ which is able to model coalition formation through argumentation. Our formalism includes two different modalities, $\langle\langle A \rangle\rangle$ and $\langle| A |\rangle$, which refer to different kinds of abilities agents may have. Note that the original operator $\langle\langle A \rangle\rangle$ is used to reason about the pure ability of the very group $A$. However, the question whether it is reasonable to assume that the members of $A$ collaborate is not taken into account in $\text{ATL}$. With the new operator $\langle| A |\rangle$ we try to close this gap, providing also a way to focus on sensible coalition structures. In this context, “sensible” refers to acceptable coalitions with respect to some argumentative semantics (as characterized in Def. 8).

Furthermore, we have defined the formal machinery required for characterizing argument-based coalition formation in terms of the proposed operator $\langle| A |\rangle$. Coalitions can be computed in terms of a given argumentation semantics, which can be given as a parameter within our model, thus providing a modular way of analyzing the results associated with different alternative semantics. This allows us to compare the ability of agents to form particular coalitions and study emerging properties regarding different semantics. Additionally, as outlined in Section 6 the model checking algorithm used in $\text{ATL}$ can be extended to $\text{CoalATL}$ by integrating suitable proof procedures for argumentation semantics. We have shown that the model checking problem is $\Delta^P_2$-complete/easy for the most natural cases.

Acknowledgements. This research was partially funded by the Projects DAAD-SeCyT (DA0609) and PGI-UNS (24/ZN10).

References


