A Finer Grained Modeling of Rational Coalitions Using Goals

Nils Bulling and Jürgen Dix
Computational Intelligence Group
Department of Computer Science – Clausthal University of Technology
38678 Clausthal-Zellerfeld – GERMANY
E-mail: {bulling,dix}@in.tu-clausthal.de

Abstract

We propose an extension of *Coalitional* ***ATL*** (a logic for reasoning about coalitions and their formation process, see [10]) by goals. This *goal framework* allows for a finer grained modeling of coalitions: *Coalitional frameworks*, based on Dungs's abstract argumentation framework, are used to point out *conflicts* between agents, and *goals* refer to agents' subjective incentives to join (or not to join) coalitions. We focus on two different aspects for cooperation allowing a more practical modeling of systems.

Keywords: Modal Logic, coalition formation, goals, game theory, multi-agent systems

1 INTRODUCTION AND MOTIVATION

In the context of multi-agent systems (**MAS**), modeling and reasoning about *coalitions* and their *abilities* are of utmost importance. In particular, it has been shown recently that *argumentation* provides a sound setting to model *reasoning about coalition formation* in multi-agent systems [3, 4] and that *alternating-time temporal logic* (**ATL**) [2] can successfully be used to model **MAS**. The latter is probably the most influential logic of strategic ability that has emerged in recent years. However, it lacks to express *why* agents should cooperate: The focus is solely on pure *abilities* of predefined coalitions (whatever the reasons for cooperation might be).

In [10] a first step in merging both areas was taken. The presented logic, *Coalitional* ***ATL*** ($\mathbf{ATL^c}$), focuses on *reasonable* coalitions. The main question addressed is the following: "Is there a reasonable coalition which can enforce φ?" rather than "Does a given group A has the ability to enforce φ?". Here, *reasonable* refers to the computation of coalitions modeled in terms of a given argumentation semantics [12] (e.g. stable or preferred) in the context of coalition formation [3].

One of the features of $\mathbf{ATL^c}$ is that it is based on a framework combining two areas: Argumentation and **ATL**. Hence, for simplicity, the coalition formation process took place in the *initial state* only and did not address the question *why* agents should cooperate but solely *why not* to cooperate. In this paper we propose a way to overcome the latter shortcoming[1]. Indeed, in [10] it was already pointed out that in order to join a coalition agents usually require some kind of *incentive* (i.e. sharing common goals or getting rewards). Here, we take this statement seriously and propose a *goal framework*, incorporated into $\mathbf{ATL^c}$ models, to allow a finer grained modeling of coalitions: *Coalitional frameworks* are used to point out *conflicts* between agents, and *goals* refer to agents' subjective incentives to join coalitions.

The rest of the paper is structured as follows. Firstly, we recall the argumentation framework used for coalition formation and *Coalitional* **ATL**. Secondly, we introduce an abstract *goal framework* and

[1] The first point was already very briefly addressed in [10]

show how it can be combined with our new logic to model agents' incentives to join coalitions in a temporal setting, besides conflicts modeled by *coalitional frameworks*. We conclude with an example and discussion of future and related work.

2 PRELIMINARIES: COALITIONS AND ATL

In the next two subsections we briefly recall the argumentative approach to coalition formation [3] and *Coalitional **ATL*** introduced in [10].

2.1 Coalitions and Argumentation

This approach to coalition formation is motivated by [3], where an argumentation framework for generating coalition structures is defined. The basic notion is that of a *coalitional framework* (**CF**)[2] modeling conflicts between agents. Such a framework is a tuple $\mathcal{CF} = (\mathbb{A}\mathrm{gt}, \mathcal{D})$ where $\mathbb{A}\mathrm{gt}$ is a non-empty set of agents[3] and $\mathcal{D} \subseteq \mathbb{A}\mathrm{gt} \times \mathbb{A}\mathrm{gt}$ is a *defeat* (or *attack*) relation[4] between agents. The set $\mathbb{CF}(\mathbb{A}\mathrm{gt})$ denotes all coalitional frameworks over $\mathbb{A}\mathrm{gt}$.

Let $\mathcal{CF} = (\mathbb{A}\mathrm{gt}, \mathcal{D})$ be a coalitional framework. For agents $a, a' \in \mathbb{A}\mathrm{gt}$, we say that a *defeats* a' iff $a\mathcal{D}a'$. The defeat relation represents conflicts between elements of $\mathbb{A}\mathrm{gt}$; for instance, two agents may rely on the same (unique) resource or they may have disagreeing goals, which prevents them from cooperation.

Defeats are defined between *single* agents. As we are interested in the formation of coalitions it is reasonable to consider conflicts between coalitions. In argumentation theory many different semantics have been proposed to define ultimately accepted arguments [12, 11]. Here arguments can be seen as agents. We apply this rich framework to provide different ways to coalition formation. Using an abstract viewpoint, a *semantics* for a coalitional framework $\mathcal{CF}$ is a(n) (isomorphism invariant) mapping $\mathfrak{sem}$ which assigns to a given coalitional framework $\mathcal{CF}$ a set of subsets of $\mathbb{A}\mathrm{gt}$, i.e., $\mathfrak{sem}(\mathcal{CF}) \subseteq \mathcal{P}(\mathbb{A}\mathrm{gt})$.

In this paper we focus on the *stable semantics* where a coalition $C \subseteq \mathbb{A}\mathrm{gt}$ is called *stable* iff it is *conflict-free* (i.e. none of its members is defeating another member) and all agents outside C are defeated by some member of C; formally, C is stable iff $\forall a, a' \in C\ \neg(a\mathcal{D}a')$ and $\forall a \in \mathbb{A}\mathrm{gt}\backslash C \exists a' \in C\ (a'\mathcal{D}a)$. For more details on various semantics we refer to [10].

Example 1 (Stable coalitions) *In Figure 1* $\{a_1, a_3\}$ *(resp.* $\{a_1, a_2\}$ *and* $\{a_1, a_2, a_3\}$*) is the unique stable coalition of coalitional framework* $\mathcal{CF}_1$ *(resp.* $\mathcal{CF}_2$ *and* $\mathcal{CF}_3$*).*

2.2 Coalitional ATL

In this section we recall *Coalitional **ATL*** (**ATL^c**) from [10].

Definition 1 ($\mathcal{L}_{ATL^c}$ [10]) *The logic* Coalitional **ATL**, $\mathcal{L}_{ATL^c}$, *is given over a finite set* $\mathbb{A}\mathrm{gt} = \{a_1, \ldots, a_k\}$ *of agents and* Π *of propositions and consists of all formulae* φ *defined by the following grammar:*

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

[2]Differently from [3, 10] we do present argumentation frameworks *without* preferences.
[3]We restrict ourselves to agents, in [3, 10] the notion was more generic.
[4]Without preferences the notions attack and defeat coincide.

$p \in \Pi$, *and* $A \subseteq \mathbb{A}\mathrm{gt}$. *A formula* φ *(resp.* γ*) is called* state *(resp.* path*)* formula*. Pure* ***ATL*** *is given by all formulae that do not contain modalities* $\langle\!| A |\!\rangle$.

Informally, $\langle\!\langle A\rangle\!\rangle\varphi$ expresses that agents A have a *collective strategy to enforce* φ. **ATL**[c] formulae include the usual temporal operators: $\bigcirc$ (*in the next state*), $\Box$ (*always from now on*) and $\mathcal{U}$ (strict *until*). Additionally, $\Diamond$ (*now or sometime in the future*) can be defined as $\Diamond\varphi \equiv \top\,\mathcal{U}\,\varphi$. The intended reading of $\langle\!| A |\!\rangle\varphi$ is that the group A of agents *is able to form a coalition* $B \subseteq \mathbb{A}\mathrm{gt}$ *such that* A *and* B *are not disjunct*, $A \cap B \neq \emptyset$, and B *can enforce* φ. Coalition formation is modeled by the formal argumentative approach in the context of coalition formation, as described in the previous section.

Coalitional **ATL** allows to reason about the ability of building coalition structures, and not only about an *a priori* specified group of agents (as it is the case for $\langle\!\langle A\rangle\!\rangle\varphi$). Compared to pure **ATL** [2], a formula like $\langle\!| A |\!\rangle\varphi$ does *not* refer to the ability of A to enforce φ, but rather to the ability of A to *constitute* a coalition B, such that $A \cap B \neq \emptyset$, and then, in a second step, to the ability of B to enforce φ.

Formulae are interpreted over extensions of concurrent game structures [2] by coalitional frameworks and its tools. A pure *concurrent game structure* (**CGS**) is given by a tuple

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

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

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

A (memoryless) *strategy* of agent a is a function $s_a : Q \rightarrow Act$ such that $s_a(q) \in d_a(q)$. We denote the set of such functions by Σ_a. A *collective strategy* s_A for team $A \subseteq \mathbb{A}\mathrm{gt}$ specifies an individual strategy for each agent $a \in A$; the set of A's collective strategies is given by $\Sigma_A = \times_{a \in A}\Sigma_a$ and $\Sigma := \Sigma_{\mathbb{A}\mathrm{gt}}$. The *outcome* of strategy s_A in state q is defined as the set of all paths that may result from executing s_A: $out(q, s_A) = \{\lambda \in \Lambda_{\mathcal{M}}(q) \mid \forall i \in \mathbb{N}_0\ \exists \vec{\alpha} = \langle \alpha_1, \ldots, \alpha_k\rangle \in d(\lambda[i])\ \forall a \in A\ (\alpha_a = s^a_A(\lambda[i]) \wedge o(\lambda[i], \vec{\alpha}) = \lambda[i+1])\}$, where s^a_A denotes agent a's part of the collective strategy s_A.

Finally, models for *Coalitional* ***ATL*** are given by **CGS**'s *extended by* $\zeta : Q \rightarrow (\mathcal{P}(\mathbb{A}\mathrm{gt}) \rightarrow \mathbb{CF}(\mathbb{A}\mathrm{gt}))$, a function which assigns a *coalitional framework* over $\mathbb{A}\mathrm{gt}$ to each state of the model subjective to a given group of agents[5], and an (argumentative) semantics $\mathfrak{sem}$.

Definition 2 (CGM) *A* coalitional game model ***CGM*** *is given by* $\mathcal{M} = \langle \mathbb{A}\mathrm{gt}, Q, \Pi, \pi, Act, d, o, \zeta, \mathfrak{sem}\rangle$ *where the first 7-tuple is a* ***CGS****,* $\zeta : Q \rightarrow (\mathcal{P}(\mathbb{A}\mathrm{gt}) \rightarrow \mathbb{CF}(\mathbb{A}\mathrm{gt}))$*, and* $\mathfrak{sem}$ *is an argumentation semantics. The set of all* ***CGM****'s over the denoted elements is defined as* $\mathbb{M}(\mathbb{A}\mathrm{gt}, Q, \Pi, Act, \mathfrak{sem}, \zeta)$*. Often, we will leave out the parameters and assume the standard element names.*

Truth of $\mathcal{L}_{ATL^c}$ formulae is defined below.

Definition 3 (ATL[c] [10]) *Let* $\mathcal{M} \in \mathbb{M}$*. Semantics of* $\varphi, \psi \in \mathcal{L}_{ATL^c}$ *is given as follows:*

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

[5] In [10] coalitional frameworks did not depend on a given group of agents.

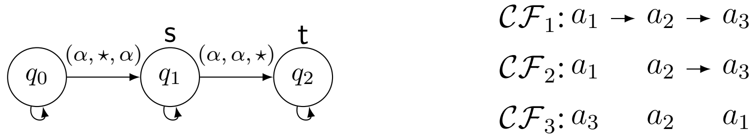

Figure 1: A simple **CGS** and three *coalitional frameworks*. There are three agents a_1, a_2, and a_3 each of them can perform action α or β; $\star$ is a placeholder for any of the two. Action profiles not mentioned explicitly cause the system to stay in the same state; for example, $o(q_1, (\beta, \alpha, \alpha)) = q_1$. In the coalitional frameworks $\mathcal{CF}_1$, $\mathcal{CF}_2$, and $\mathcal{CF}_3$ attacks are represented by arcs; for instance, agent a_2 attacks a_3 in $\mathcal{CF}_1$.

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

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

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

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

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

$\mathcal{M}, q \models \langle\!|A|\!\rangle \gamma$ *iff there is a* valid coalition B *wrt* A *and* q *such that* $\mathcal{M}, q \models \langle\langle B\rangle\rangle \gamma$.

where B is called *valid*[6] if, and only if, $B \in \mathfrak{sem}(\zeta(q, A))$ and $A \cap B \neq \emptyset$.

Example 2 (Simple CGM) *We consider the* ***CGS*** *shown in Figure 1. We can extend the model to a* ***CGM*** $\mathcal{M}$ *as described in the following. We set* $\zeta(q_0) = \mathcal{CF}_3$ *and* $\zeta(q) = \mathcal{CF}_2$ *for* $q = q_1, q_2$ *and use the stable semantics as argumentation semantics. Then we have, for instance, that* $\mathcal{M}, q_0 \models \neg\langle\langle X\rangle\rangle \Diamond \mathsf{t}$ *for* $X \subsetneq \mathbb{A}\mathrm{gt}$ *and* $\mathcal{M}, q_0 \models \langle\langle \mathbb{A}\mathrm{gt}\rangle\rangle \Diamond \mathsf{t}$*; hence, also* $\mathcal{M}, q_0 \models \langle\!|a_2|\!\rangle \Diamond \mathsf{t}$ *since* $\mathbb{A}\mathrm{gt}$ *is a stable coalition in* q_0.

3 COOPERATION OF AGENTS WITH GOALS

Why should agents join coalitions? They must have reasons to do so. In this paper, 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 **ATL^c**, discuss its benefits and illustrate it with an example.

3.1 Goals and Agents

Pro-activeness and *social ability* are among the widely accepted characteristics of intelligent agents [15]. In BDI frameworks, *goals* (or *desires*) and *beliefs* play an important role [9, 14].

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.

[6]The original definition in [10] was slightly different. Firstly, we assumed that A is always a valid coalition (i.e.$B \in \mathfrak{sem}(\zeta(q, A)) \cup \{A\}$); and secondly, as mentioned above the function ζ was agents independent (i.e. $\zeta : Q \to \mathbb{CF}(\mathbb{A}\mathrm{gt})$). We come back to these modification in Section 3.3.

Definition 4 ($\mathcal{G}$, goal mapping $\mathfrak{g}$) *Let $\mathcal{G}_a$ be a non-empty set of elements (set of* goals*), one for each agent $a \in \mathbb{A}\text{gt}$, and $\mathcal{G} := \bigcup_{a\in\mathbb{A}\text{gt}} \mathcal{G}_a$. By "$g$" we denote a typical element from $\mathcal{G}$. A* goal mapping *is a function $\mathfrak{g} : \mathbb{A}\text{gt} \to (Q^+ \to \mathcal{P}(\mathcal{G}))$ 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 $\Diamond \mathsf{s}$ 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 5 (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 $\mathfrak{g}_a$ for $a \in \mathbb{A}\text{gt}$ 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, $o \hookrightarrow g$, 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, $o \hookrightarrow^s g$; 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 $\Box \mathsf{t}$ satisfies goal $\Box(\mathsf{t} \vee \mathsf{s})$ an supports goal $\Diamond \mathsf{t}$.

3.2 Specifying Goals and Objectives

In this section, we propose **ATL** (resp. **ATL$^\mathbf{c}$**) as a *goal specification language*. 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 [5, 8, 7].

Goals *formulated in **LTL*** are very intuitive. Formulae like $\Diamond \mathsf{rich}$ (*eventually being rich*), $\bigcirc \mathsf{takeUmbrella}$ (*take umbrella in the next moment*), or $\Box\Diamond \mathsf{sleep}$ (*going to sleep again and again*) have clear interpretations. But goals *formulated in **CTL*** can be ambiguous. A goal like $\mathsf{A}\Diamond \mathsf{rich}$[7] does not seem fundamentally different from $\Diamond \mathsf{rich}$ 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: $\neg\langle\langle b\rangle\rangle\Diamond \mathsf{critical}$. However, it might be acceptable for a that b together with another agent c enters the

[7] The operator A refers to all possible paths starting in a state. In **ATL** this operator can be expressed as $\langle\langle\emptyset\rangle\rangle$.

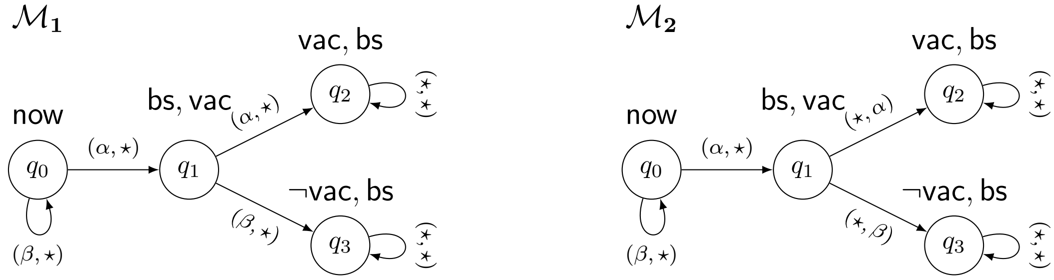

Figure 2: Two simple models showing that **ATL** goals are useful. $\star \in \{\alpha, \beta\}$ is used as a placeholder for any of the two actions.

critical section because then c has to unlock resources a could use instead. Let us consider a more detailled example.

Example 3 (ATL-goals) *In the example we consider two agents a and b. Both agents can perform actions α and β. The first agent, leader of a research group, would like to get a better salary (*bs*) and wants to retain the power to decide when to take vacation (*vac*). So, a's goal can be expressed as $\gamma \equiv \Box(\neg\mathsf{now} \rightarrow \mathsf{bs} \wedge \mathsf{vac})$. Interpreting the models shown in Figure 2 purely temporal (i.e. without action profiles) the **CTL** formula $\mathsf{E}\gamma$ is satisfied in q_0 in both models: There are q_0-paths which satisfy γ. On the other hand, $\mathsf{A}\gamma$ 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 $\gamma' \equiv \Box(\neg\mathsf{now} \rightarrow \mathsf{bs} \wedge \neg\langle\langle b\rangle\rangle\Diamond\neg\mathsf{vac})$, or equivalently in this example $\Box(\neg\mathsf{now} \rightarrow \mathsf{bs} \wedge \langle\langle a\rangle\rangle\Box\mathsf{vac})$. Now, it is easy to see that $\mathcal{M}_1, q_0 \models \langle\langle a\rangle\rangle\gamma'$ but $\mathcal{M}_2, q_0 \not\models \langle\langle a\rangle\rangle\gamma'$. 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 6 (ATL-Goal) *Let γ, γ' be **ATL** path formulae. An* **ATL**-goal *has the form γ or $\gamma \wedge \gamma'$*[8].

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 **ATL^c** formula $\langle\!|A|\!\rangle\gamma$. The question is whether there is a *rational* group to bring about γ; thus, only agents which gain advantage when γ is fulfilled should cooperate. Hence, we consider γ as objective; so, in general all Coalitional **ATL** path formulae.

Definition 7 (ATL^c -objective) *An* **ATL^c** -objective *is an $\mathcal{L}_{ATL^c}$* path *formula.*

3.3 Goals as a Means for Cooperation

In this section we link together Coalitional **ATL** with the goal framework described above. The syntax of the logic is given as in Section 2.2. The necessary change takes place in the semantics. We redefine what it means for a coalition to be *valid*.

[8]Note that $\gamma \wedge \gamma'$ is not an **ATL** path formula anymore.

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. ***CGM's with goals*** are given as a straightforward extension of **CGM**'s (cf. Definition 2).

Definition 8 (CGM with goals) *A* **CGM** with goals *(**CGMg**) $\mathcal{M}$ is given by a model of* $\mathbb{M}(Q, \mathbb{A}\text{gt}, \Pi, \mathfrak{sem}, \zeta)$ *extended by a set of goals $\mathcal{G}$ and a goal mapping $\mathfrak{g}$ over $\mathcal{G}$. The set of all such models is denoted* $\mathbb{M}^g(Q, \mathbb{A}\text{gt}, \Pi, \mathfrak{sem}, \zeta, \mathcal{G}, \mathfrak{g})$ *or just* $\mathbb{M}^g$ *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 $i,j \in \mathbb{N}_0 \cup \{\infty\}$ and $i \leq j$. 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 ε for $i < 0$ where $i \in \mathbb{Z} \cup \{\infty\}$. Furthermore, given a history h and a path or history λ the combined path/history starting with h extended by λ is denoted by $h \circ \lambda$.

Finally, we present the semantics of **ATL^c** *with* goals. It is similar to Definition 3. Here, however, it is necessary to keep track of the steps, visited states, made to determine the goals of the agents.

Definition 9 (Semantics of $\mathcal{L}^{vc^g}_{ATL}$) *Let $\mathcal{M}$ be a **CGMg**, q a state, φ, ψ state-, γ a path formula, and $i,j \in \mathbb{N}_0$. Semantics to $\mathcal{L}^{vc^g}_{ATL}$ formulae is given as follows:*

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

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

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

$\mathcal{M}, q, \tau \models \langle\langle A\rangle\rangle\gamma$ *iff there is a strategy $s_A \in \Sigma_A$ such that for all $\lambda \in out(q, s_A)$ it holds that* $\mathcal{M}, \lambda, \tau \models \varphi$

$\mathcal{M}, q, \tau \models \langle\!| A |\!\rangle\gamma$ *iff there is $A' \in$ **VC**$^g(q, A, \gamma, \tau)$ such that* $\mathcal{M}, q, \tau \models \langle\langle A\rangle\rangle\gamma$

$\mathcal{M}, \lambda, \tau \models \varphi$ *iff* $\mathcal{M}, \lambda[0], \tau \models \varphi$

$\mathcal{M}, \lambda, \tau \models \Box\varphi$ *iff for all i it holds that* $\mathcal{M}, \lambda[i], \tau \circ \lambda[1,i] \models \varphi$

$\mathcal{M}, \lambda, \tau \models \bigcirc\varphi$ *iff it holds that* $\mathcal{M}, \lambda[1], \tau \circ \lambda[1] \models \varphi$

$\mathcal{M}, \lambda, \tau \models \varphi\,\mathcal{U}\,\psi$ *iff there is a j such that $\mathcal{M}, \lambda[j], \tau \circ \lambda[1,j] \models \psi$ and for all $0 \leq i < j$ it holds that* $\mathcal{M}, \lambda[i], \tau \circ \lambda[1,i] \models \varphi$.

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

All the new functionality provided by goals is captured by the new valid coalition function **VC**^{g}

Definition 10 (Valid coalitions,VC$^g(q, A, o, \tau)$) *Let $\mathcal{M} \in \mathbb{M}^g$, $\tau \in Q^+$, $A, B \subseteq \mathbb{A}\text{gt}$, o an **ATL^c** objective.*

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

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

2. *there are goals $g_{b_i} \in \mathfrak{g}_{b_i}(\tau)$, one per agent $b_i \in B$, such that $o \hookrightarrow_{\mathcal{M},\tau,B} g_{b_1} \wedge \cdots \wedge g_{b_{|B|}}$*

The set $\mathbf{VC}^g(q, A, o, \tau)$ 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 ζ and a sequence of states τ are required. The intuition of τ is that it represents the history (the sequence of states visited so far including the current state). So, τ is used to determine which goals of the agents are still active.

The main semantical differences from the version defined in [10] (cf. Definition 3) are as follows:

1. Goals are used to rule out an agent's membership in coalitions.
2. Coalitional frameworks are assigned to states *and depend on* groups of agents.
3. It is not required that a forming coalition A, specified by $\mathbf{VC}^g(\dots, A, \dots)$, is completely contained in B in order to validate B: We just assume that the formed coalition should contain *some* member of A.
4. The forming coalition A is not valid by definition.

Note that the last point is a generalization. The "old" operator $\langle\!|A|\!\rangle^{\text{old}}\gamma$ as defined in [10] which assumes that A is always an acceptable coalition can easily be expressed by $\langle\!|A|\!\rangle^{\text{new}}\gamma \vee \langle\!\langle A \rangle\!\rangle\gamma$ where $\langle\!|A|\!\rangle^{\text{new}}$ is equal to the former but the new version given in Definition 3 (cf. Definition 9 for a path-based semantics) of valid coalition is used. Hence, A is only a possible candidate coalition if it is actually acceptable according to the argumentation semantics.

Finally, 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 11 (Satisfaction of goals) *Let g be an **ATL**-goal, o an $\mathcal{L}_{ATL}^{vc^g}$-objective, and $\tau \in Q^+$. We say that objective o* satisfies *g, for short $o \hookrightarrow_{\mathcal{M},\tau,B} g$, with respect to $\mathcal{M}, \tau$, and B if, and only if, there is a strategy $s_B \in \Sigma_B$ such that*

1. *for all $\lambda \in out(\tau[\infty], s_B)$ it holds that $\mathcal{M}, \lambda, \tau \models o$ implies $\mathcal{M}, \lambda \models g$, and*
2. *that there is some path $\lambda \in out(\tau[\infty], s_B)$ 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 Defintion 9.

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 without the $\langle\!|A|\!\rangle$-rule) by the following semantic rule: $\mathcal{M}, \lambda \models \gamma_1 \wedge \cdots \wedge \gamma_n$ iff $\mathcal{M}, \lambda \models \gamma_i$ for $i = 1, \dots, n$.

Coming back to the satisfaction of goals. Let $\langle\!|A|\!\rangle'$ denote a coalitional operator with semantics equal to $\langle\!|A|\!\rangle$ but without the second condition in Definition 11 (note that this definition is needed to determine the valid coalitions).

Proposition 1 *Let γ be an **ATL**^{c} path formula, $\mathcal{M} \in \mathbb{M}^g$, $q \in Q_{\mathcal{M}}$, $A \subseteq \mathbb{A}\text{gt}$, and $\tau \in Q^+$. Then we have that $\mathcal{M}, q, \tau \models \langle\!|A|\!\rangle\gamma$ if, and only if, $\mathcal{M}, q, \tau \models \langle\!|A|\!\rangle'\gamma$*

Proof. The proof is simple. Firstly, observe that $\forall \lambda \in out(q, s_B)\ (\lambda \models \gamma)$ implies $\exists \lambda \in out(q, s_B)\ (\lambda \models \gamma)$ since the outcome is never empty. Then we have $q, \tau \models \langle\!|A|\!\rangle\gamma$ iff $\exists B \in \mathbf{VC}^g\ \exists s_B \in \Sigma_B\ \forall \lambda \in out(q, s_B)\ (\lambda, \tau \models \gamma)$ iff $\exists B \in \mathbf{VC}^g\ \exists s_B \in \Sigma_B\ (\forall \lambda \in out(q, s_B)\ (\lambda, \tau \models \gamma)$ and $\exists \lambda \in out(q, s_B)\ (\lambda, \tau \models \gamma))$ $(\star)$ iff $\exists B \in \mathbf{VC}^{g'}\ (q, \tau \models \langle\!\langle B\rangle\!\rangle\gamma)$ iff $q, \tau \models \langle\!|A|\!\rangle'\gamma$ where $\mathbf{VC}^{g'}$ is equal to $\mathbf{VC}^g$ except for condition 2 in Definition 11. Finally, it is easy to observe that the equivalence $(\star)$ holds since we can just take s_B as goal satisfaction strategy. (We omitted the model and the parameters of $\mathbf{VC}^g$ ad $\mathbf{VC}^{g'}$.) ■

Remark 2 (Preferred goals) *In the abstract goal framework presented in Section 3.1 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 a 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.

3.4 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 $\Diamond\mathsf{p}$ 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 $\Diamond\mathsf{p}$.) 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 $\Diamond\mathsf{p} \wedge \Box\mathsf{q}$ should be specialized to $\Box\mathsf{q}$ if a state is reached in which p holds. In [6] such a progression procedure is presented for first-order **LTL**.

4 AN EXAMPLE

In our example we consider a simple (political) coalition formation scenario. The scenario consists of five agents $\mathbb{A}\mathrm{gt} = \{a_1, a_2, a_3, a_4, a_5\}$. Agents one to four try to form a coalition and agent a_5 has the role of an independent entity controlling the process. He has to agree with any formed coalition before it can be constituted. The scenario is modeled in Figure 3 and is explained below. Arcs are labeled with coalitions which can enforce the very transition. This can easily be transformed into the

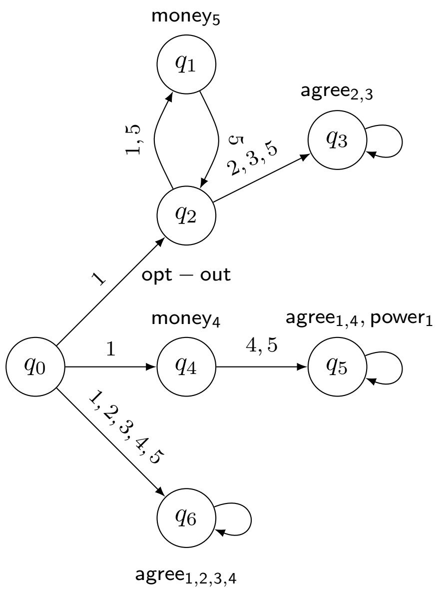

Figure 3: A **CGS** modeling a coalition formation scenario.

standard representation via action profiles. For instance, an arc labeled with $\{a_1, a_2\}$ corresponds to an action profile $\langle \alpha, \beta, \star, \star, \star \rangle$ depending only on the choices of agents a_1 and a_2. Next we define how to determine acceptable coalitions; that is, coalitions which can in principle be formed. This is modeled by coalitional frameworks. For simplicity, and due to the reason that our focus is on goals (the new concept in this paper), we assume coalitions which do not contain a subgroup consisting of a_1 together with a_2 or a_3 at the same time are acceptable; that is, for all states q and all coalitions A we set $\mathfrak{sem}(\zeta(q)(A)) = \{C \in \mathcal{P}(\mathbb{A}\mathrm{gt}) \mid \{a_1, a_2\}, \{a_1, a_3\} \not\subseteq C\}$.

The story behind the picture is as follows (where we are particularly interested in agent a_1). Agent a_1's main interest is to have power and to decide on it on his own, this is modeled by $\Diamond\langle\langle 1 \rangle\rangle\Box\mathsf{power}_1$: Sometime in the future agent a_1 wants to ensure (on his own) to stay in power forever. Besides that (if his primary goal can for instance not be fulfilled) the agent does not want that agents a_2 and a_3 have the ability to form a coalition on their own (permitted by a_5). This is also reflected in the coalitional frameworks. Intuitively, a_1 and a_2 (resp. a_3) are political opponents and cannot agree on similar goals. The latter goal can be expressed as **ATL** path formula: $\Box\neg\langle\langle a_2, a_3 \rangle\rangle\Diamond\mathsf{agree}_{2,3}$. In principal, all agents can agree to achieve an agreement: $q_0 \models \langle\langle \mathbb{A}\mathrm{gt} \rangle\rangle\bigcirc\mathsf{agree}_{1,2,3,4}$. However, this coalition is not rational due to the coalitional framework which prohibits a_1 to be in the same coalition as a_2 and a_3. Thus, we have that $q_0 \not\models \llparenthesis X \rrparenthesis\bigcirc\mathsf{agree}_{1,2,3,4}$ for all $X \subseteq \mathbb{A}\mathrm{gt}$.

In state q_0 agent a_1 can decide to opt out of the coalition formation process; however, this opens the possibility for $\{a_2, a_3\}$ to enforce $\mathsf{agree}_{2,3}$ (without good reason a_5 will not interfere). Thus, we have $q_2 \models \langle\langle a_2, a_3, a_5 \rangle\rangle\bigcirc\mathsf{agree}_{2,3}$, $q_2 \models \llparenthesis a_2 \rrparenthesis\bigcirc\mathsf{agree}_{2,3}$, and $q_2 \models \llparenthesis a_3 \rrparenthesis\bigcirc\mathsf{agree}_{2,3}$ if we assume that a_2 and a_3 have, for instance, the goal $\Diamond\mathsf{agree}_{2,3}$. The unique path satisfing the first three formulae does also satisfy the latter goal. However, a_1 still has the option to convince a_5 (by giving money) to veto coalition $\{a_2, a_3\}$, thus $q_0 \models \llparenthesis a_1 \rrparenthesis\Box\neg\mathsf{agree}_{2,3}$ assuming that agent a_5 likes to get money ($\Diamond\mathsf{money}_5$). Finally, for the same reason a_1 (resp. a_4) together with a_5 are able to form a coalition which prevents a_2 or a_3 to reach an agreement on their own: $q_0 \models \llparenthesis X \rrparenthesis\Diamond\mathsf{agree}_{1,4}$ where $X \in \{\{a_1, a_5\}, \{a_4, a_5\}\}$. Agent a_4 needs additional money to be able to take part in a coalition.

5 RELATED WORK, CONCLUSION

Related Work. Quantified Coalition Logic (QCL), proposed in [1], follows an idea similar to our proposal but the focus is different: While we aim at *rational* coalitions, in [1] the focus is on succinctness of formulae.

The authors introduce a new operator $\langle P \rangle$ which is read as follows: *"there exists a coalition C satisfying property P such that C can achieve ϕ"*. However, P describes purely set theoretic properties (e.g. $A \subset B$) and the quantification takes place in modal operators (as in $\langle\!\langle A \rangle\!\rangle$). In addition to that, QCL is *no* more expressive than *coalition logic* [13] where our proposal extends **ATL** by *new concepts*.

Indeed, quantification in the objective language is mentioned in the QCL paper, too (and also served as motivation for this paper). But it is not further pursued because of the high computational complexity. We believe, however, that quantification in the object language is justifiable by its additional expressive power: Complexity is always a tradeoff wrt expressiveness. The syntax can often be suitably restricted (number of coalitional variables, nestings of quantifiers and so forth) to get better computational complexity.

Conclusion. We have introduced goals to the logic Coalitional **ATL**, which allows to model the coalition formation process. In the goal framework we proposed **ATL** as a suitable language to formulate them. Finally, these goals can be implemented using Coalitional **ATL**.

We did not comment on complexity issues wrt model checking of our logic. Also the interplay with fixed points operators seems to be an interesting issue for further research.

REFERENCES

[1] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. Quantified coalition logic. In *IJCAI*, pages 1181–1186, 2007.

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

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

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

[5] Fahiem Bacchus and Froduald Kabanza. Planning for temporally extended goals. *Ann. Math. Artif. Intell.*, 22(1-2):5–27, 1998.

[6] Fahiem Bacchus and Froduald Kabanza. Using temporal logics to express search control knowledge for planning. *Artificial Intelligence*, 116(1-2):123–191, 2000.

[7] Chitta Baral, Vladik Kreinovich, S. Sarkar, and Raúl Trejo. Computational complexity of planning with temporal goals. In *IJCAI*, pages 509–514, 2001.

[8] Chitta Baral and Jicheng Zhao. Non-monotonic temporal logics for goal specification. In *IJCAI*, pages 236–242, 2007.

[9] M.E. Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987.

[10] N. Bulling, C. Chesñevar, and J. Dix. Modelling coalitions: ATL+argumentation. In *Proceedings of AAMAS'08*, Estoril, Portugal, May 2008. ACM Press.

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

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

[13] M. Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[14] A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484, 1991.

[15] M. Wooldridge. *An Introduction to Multi Agent Systems*. John Wiley & Sons, 2002.