Multi-Agency Is Coordination And (Limited) Communication

Piotr Kaźmierczak[1,2], Thomas Ågotnes[2], and Wojciech Jamroga[3,4]

[1] Department of Computing, Mathematics and Physics, Bergen University College
[2] Department of Information Science and Media Studies, University of Bergen
[3] Computer Science and Communication, and Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg
[4] Institute of Computer Science, Polish Academy of Sciences
phk@hib.no, thomas.agotnes@infomedia.uib.no, w.jamroga@ipipan.waw.pl

Abstract. Systems within the agent-oriented paradigm range from ones where a single agent is coupled with an environment to ones inhabited by a large number of autonomous entities. In this paper, we look at what distinguishes single-agent systems from multi-agent systems. We claim that *multi*-agency implies limited coordination, in terms of action and/or information. If a team is characterized by full coordination both on the level of action choice and the available information, then we may as well see the team as a single agent in disguise. To back the claim formally, we consider a variant of Alternating-time Temporal Logic ATL where each coalition operates with a single indistinguishability relation. For this variant, we propose a truth-preserving translation of formulas and models in the syntactic fragment of ATL where only singleton coalitions are allowed. In consequence, we show that assuming unified view of the world on part of each coalition reduces the full language of ATL to its single-agent fragment when a model is given.

1 Introduction

Agent-based models become a suitable foundation for IT environments nowadays. More and more systems involve social as much as technological aspects, and even those that focus on technology are often based on distributed components exhibiting self-interested, goal-directed behavior. Moreover, the components act in environments characterized by incomplete information and uncertainty. The field of *multi-agent systems* [28] studies the whole spectrum of phenomena ranging from agent architectures to communication and coordination in agent groups to agent-oriented software engineering. The theoretical foundations are mainly based on game theory and formal logic.

Systems within the agent-oriented paradigm display various degrees of multiplicity, from systems where a single agent is coupled with an environment (often used e.g. in agent-oriented programming), to massively populous ones used e.g. for agent-based simulation. What distinguishes a single-agent system from a *multi*-agent system is an interesting question in itself. In particular, is it enough that a system consists of multiple *modules* to call it multi-agent? What about semi-autonomous entities that act according to "orders" dispatched from a central unit? Or entities that act autonomously but they pursue a common goal, and act according to a joint plan? All these cases clearly display different degrees of autonomy and agency.

In this paper, we claim that multi-agency implies limited coordination, in terms of action and/or information. That is, different agents may collaborate, but they are inherently separated: each agent is individually responsible for executing his/her actions, and does that based on his/her individual view of the situation. Putting it in another way, if a team is characterized by full coordination both on the level of action/strategy choice and the available information, then we may as well see the team as a single (though composite) agent.

To back the claim formally, we use Alternating-time Temporal Logic ATL [4, 5] that combines elements of game theory, temporal logic, and epistemic logic in a neat formal framework. Coordination of coalitional strategies is implicitly given "for free" in the semantics of ATL, but the logic has many semantic variants for reasoning about coalitional play under different patterns of uncertainty. We consider a variant of ATL where each coalition operates by definition with a single indistinguishability relation (e.g., the distributed knowledge relation). For this variant, we propose a truth-preserving translation of formulas and models in the syntactic fragment of ATL where only singleton coalitions are allowed. In consequence, we show that assuming unified view of the world on part of each coalition reduces the full language of ATL to its single-agent fragment, at least when a model is given.

The main purpose of this study is philosophical. We want to understand the different degrees of autonomy and agency that arise in complex systems. Still, the reduction that we propose can be potentially used to implement model checking for some interesting semantic variants of ATL.

We begin by introducing the relevant syntactic and semantic variants of ATL in Sections 2 and 3. Then, we present our main result in Section 4. We conclude with some final remarks in Section 5.

Related Work. ATL has been studied extensively in the last 15 years. The research can be roughly divided into the computational and conceptual strands. The conceptual strand focuses on looking for the "right" semantics of ability, especially in the presence of imperfect or incomplete information. ATL has been combined with epistemic logic [24, 25, 1, 16], and several semantic variants were defined that match various possible interpretations of strategic ability [22, 18, 16]. Multiple extensions have been considered, e.g., with explicit reasoning about strategies, rationality assumptions and solution concepts [26, 23, 27, 9], agents with bounded resources [3, 8], coalition formation and negotiation [7], opponent modeling and action in stochastic environments [15, 21] and reasoning about irrevocable plans and interplay between strategies of different agents [2, 6]. Besides providing a palette of different formal interpretations for the concept of strategic ability, the research brought benefits in analysis of related verification problems, such as module checking [17].

In this paper, we are especially interested in works that redefine the "uniformity" conditions for coalitional play, based on a *single* epistemic relation for the whole coalition [13, 11, 14, 10]. Philosophically, this amounts to assuming members of the coalition to share their knowledge (or, conversely, propagate their uncertainty) at each step while executing a joint strategy. We show that – at least in the context of model checking – such a coalition can be seen as a single agent executing compound actions *in unison*.

2 Reasoning About Abilities of Agents and Coalitions

Alternating-time Temporal Logic ATL [4, 5] is a non-normal modal logic that allows for expressing properties of multi-agent systems. Specification in ATL is usually based on formulae of type $\langle\langle A\rangle\rangle\varphi$, expressing that the group of agents A has a strategy to enforce the temporal property φ no matter what the other agents do. Formulae of ATL are interpreted in *concurrent game structures* that assume synchronous execution of actions from all the agents in the system. We begin by defining the models formally. Then, we present the syntax and the semantic clauses for relevant variants of the logic.

2.1 Models: Concurrent Game Structures

In the most general case, formulas of ATL are interpreted over *imperfect information concurrent game structures* (ICGS), defined as follows:

Definition 1 (Imperfect Information Concurrent Game Structures [22]). *An* ICGS *is an 8-tuple* $M = \langle \Sigma, \Pi, Q, C, d, \delta, \sim, \pi\rangle$*, where:*

- Σ *is a finite set of agents;*
- Π *is a finite set of propositional letters;*
- Q *is a finite set of states;*
- C *is a finite set of choices/actions;*
- $d : Q\times\Sigma \to \wp(C)$ *is a* guard function *that specifies which actions are enabled for whom and where. It is assumed that in any state at least one choice must be enabled for each agent. We will usually write* $d_a(q)$ *instead of* $d(q, a)$*;*
- $\delta : Q \times C^{\Sigma} \to Q$ *is a deterministic transition function;*
- $\sim: \Sigma \to \wp(Q \times Q)$ *is a family of equivalence relations (one per agent) indicating states that are indistinguishable from agents' perspectives, and*
- $\pi : Q \to \wp(\Pi)$ *is a valuation of atomic propositions.*

Additionally, it is usually assumed that the ICGS *assigns the same sets of choices to indistinguishable states; formally: if* $q \sim_a q'$ *then* $d_a(q) = d_a(q')$*.*

We illustrate how ICGS are used to model multi-agent systems on the following example.

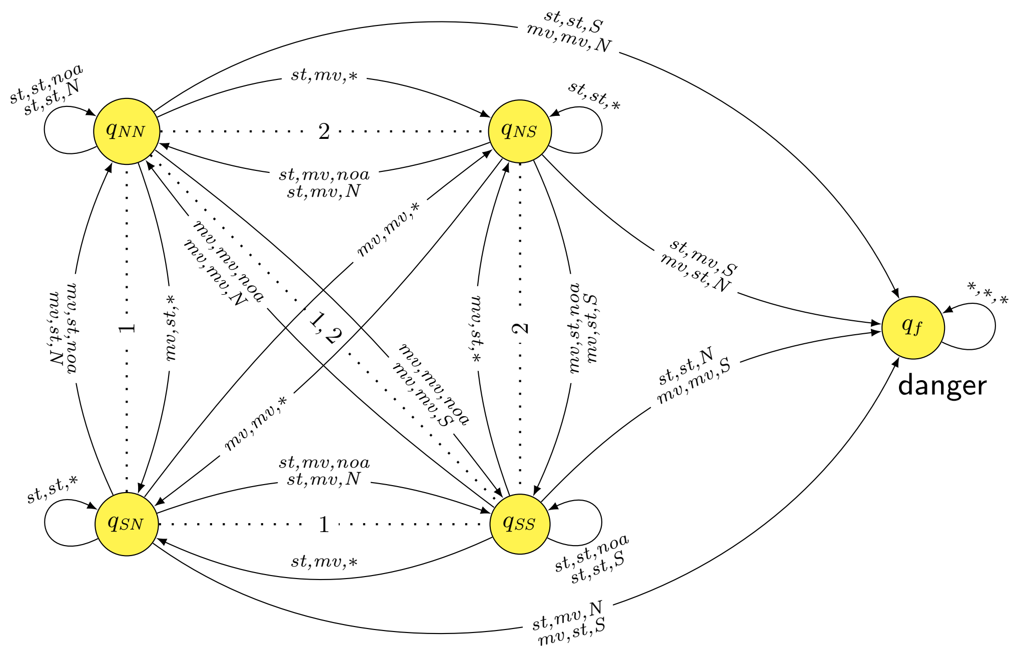

Fig. 1. A simple model of coordinated defense (M_1). The wildcard ($*$) matches any action of the appropriate player

Example 1 (Coordinated Defense). Two guards (agents 1 and 2) are supposed to protect a sensitive area from attack. They conduct surveillance of the area in parallel, from two separate locations. These can be different floors in a building, or different hills giving view to a military zone, etc. At any moment, each guard is in a position that allows him to protect either the North or the South entry to the area, but not both at the same time. Moreover, a guard can stay in the same place (action *st*) or move to the other side of the surveillance area (action *mv*). However, the landscape is confusing and the guards are no experts in reading landscape signs; in consequence, both entries and surveillance points look the same to each guard. On the other hand, guard 1 can recognize when he is in the North position and guard 2 is in the South position, because only then he can see the light from the other guard's torch. Likewise, guard 2 can only distinguish the situation when he is in the North and the other guard is in the South.

The attack – executed by the third agent a, the "attacker" – can be conducted either from the North (action N) or from the South (action S). The attacker can also refrain from attacking (action *noa*). The attack is only successful if it targets a position which, in the very next moment, will not be protected by any of the guards. In such case, the system proceeds to the "failure" state q_f, labeled by the atomic proposition danger.

A simple model of the scenario is depicted in Figure 1. The set of players is $\Sigma = \{1, 2, a\}$. States, transitions (represented by solid arrows), indistinguishability relations (represented by dotted lines), and valuation of atomic propositions can be easily read off the picture.

2.2 Syntax: ATL and Single-Agent ATL

The language of alternating-time temporal logic, formally referred to as $\mathcal{L}_{\text{ATL}}$, is defined by the following grammar:

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

where p is a propositional symbol and A is a subset of agents (called sometimes a *coalition*). We will write $\langle\langle a_1, a_2, \dots\rangle\rangle$ instead of $\langle\langle \{a_1, a_2, \dots\}\rangle\rangle$. The temporal operator $\bigcirc$ stands for "in the next moment", $\Box$ for "always from now on", and $\mathcal{U}$ for "strong until". We use the usual abbreviations of Boolean operators, plus the standard abbreviation for "sometime in the future": $\Diamond\varphi \equiv \top\,\mathcal{U}\varphi$.

We also define a syntactic fragment of ATL called the *single-agent* ATL, that allows only singleton coalitions in the formulae. The fragment, to which we will refer as $\mathcal{L}_{\text{1ATL}}$, is formally defined as follows:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \langle\langle a\rangle\rangle \bigcirc \varphi \mid \langle\langle a\rangle\rangle \Box \varphi \mid \langle\langle a\rangle\rangle \varphi \,\mathcal{U}\, \varphi$$

where p is a propositional symbol and a is an agent.

Example 2. The following formulae specify possible requirements on the Coordinated Defense scenario: $\langle\langle 1, 2\rangle\rangle \Box \neg \mathsf{danger}$ (the guards can effectively protect the system from attacks forever), $\neg\langle\langle 1, 2\rangle\rangle \Diamond \langle\langle 1, a\rangle\rangle \bigcirc \mathsf{danger}$ (the guards cannot compromise the system in such a way that, at some future moment, guard 1 can collude with the attacker for a successful attack).

2.3 Semantic Variants of ATL

Semantic variants of ATL are usually derived from different assumptions about agents' capabilities. Can the agents "see" the current state of the system, or only a part of it? Can they memorize the whole history of observations in the game? Different answers to these questions induce different semantics of strategic ability. In this section, we recall the "canonical" variants as proposed in [22]. There, a taxonomy of four strategy types was introduced and labeled as follows: I (resp. i) stands for *perfect* (resp. *imperfect*) *information*, and R (resp. r) refers to *perfect recall* (resp. *no recall*). In essence, the semantics of ATL in [22] is parameterized with the strategy type – yielding four different semantic variants of the logic, labeled accordingly (ATL_{IR}, ATL_{Ir}, ATL_{iR}, and ATL_{ir}).

The following types of strategies are used in the respective semantic variants:

– Ir: $f_a : St \to Act$ such that $f_a(q) \in d(a, q)$ for all q;
– IR: $f_a : St^+ \to Act$ such that $f_a(h) \in d(a, q_n)$ for all $h = q_0, \ldots, q_n$;
– ir: like Ir, with the additional constraint that $q \sim_a q'$ implies $f_a(q) = f_a(q')$;
– iR: like IR, with the additional constraint that $h \approx_a h'$ implies $f_a(h) = f_a(h')$, where $h \approx_a h'$ iff $h[i] \sim_a h'[i]$ for all i.

That is, strategy f_a is a conditional plan that specifies agent a's actions in each state of the system (for memoryless agents) or for every possible history of the system evolution (for agents with perfect recall). Moreover, imperfect information strategies specify the same choices for indistinguishable states (resp. histories). Finally, a collective xy-strategy F_A for a group of agents $A \subseteq \Sigma$ is a tuple of xy-strategies $(f_a)_{a \in A}$, one for each agent.

A *computation* is an infinite sequence of states $\lambda = q_0, q_1, \ldots$, and we say it is an *outcome* of strategy F_A from state q if $q_0 = q$ and for each $i > 0$, there are $c_a \in C$ choices for $a \in \Sigma \setminus A$, such that $q_{i+1} = \delta(q_i, \mathbf{c})$ where $\mathbf{c}_a = f_a([q_i]_a)$ if $a \in A$, and $\mathbf{c}_a = c_a$ for the opponents. Outcomes are therefore computations that start from the given state and follow the strategy. The set of all the outcome paths of strategy F_A from state q on is denoted by $out(q, F_A)$.

Given an ICGS M, state q and formula φ, we interpret the formula as follows:

– $M, q \models_{xy} \langle\langle A\rangle\rangle \bigcirc \varphi$ iff there exists an xy-strategy F_A for A such that, for every q' with $q \sim_A q'$, and every $\lambda \in out(q', F_A)$, we have $\lambda[1] \models_{xy} \varphi$.
– $M, q \models_{xy} \langle\langle A\rangle\rangle \Box \varphi$ iff there exists an xy-strategy F_A for A such that, for every q' with $q \sim_A q'$, and every $\lambda \in out(q', F_A)$, we have $\lambda[i] \models_{xy} \varphi$ for every position $i \geq 0$.
– $M, q \models_{xy} \langle\langle A\rangle\rangle \varphi_1 \mathcal{U} \varphi_2$ iff there exists an xy-strategy F_A such that, for every q' with $q \sim_A q'$, and every $\lambda \in out(q', F_A)$, there is a position $i \geq 0$ such that $\lambda[i] \models_{xy} \varphi_2$, and for all positions $0 \leq j < i$ we have $\lambda[j] \models_{xy} \varphi_1$.

In the above clauses, one element is not properly defined yet – namely, the coalitional indistinguishability relation $\sim_A$. Epistemic logic suggests several "canonical" ways in which collective indistinguishability can be constructed. The epistemic relation for *"everybody knows"* is defined as the union of individual relations: $\sim_A^E = \bigcup_{a \in A} \sim_a$. The relation for *common knowledge* ($\sim_A^C$) is the transitive closure of $\sim_A^E$. Furthermore, the epistemic relation for *distributed knowledge* is defined as the intersection of individual relations: $\sim_A^D = \bigcap_{a \in A} \sim_a$. Since we focus on coalitions that can freely communicate and exchange information, we assume that $\sim_A = \sim_A^D$. Notice that the distinction is not relevant for the main result in this paper, which proceeds by embedding coalition-uniform abilities in 1ATL. This is because, for individual agents, $\sim_{\{a\}}^D = \sim_{\{a\}}^E = \sim_{\{a\}}^C = \sim_a$.

We illustrate how the semantics works on the Coordinated Defense example.

Example 3. For model M_1 from Example 1 and formulae from Example 2 we have the following. $M_1, q_{SN} \models_{ir} \langle\langle 1, 2\rangle\rangle \Box \neg \mathsf{danger}$ because (i) the coalition $\{1, 2\}$ has distributed knowledge that the initial state is precisely q_{SN} and (ii) from q_{SN}, the memoryless strategy where each guard does *st* in every state avoids q_f no matter what the attacker does. On the other hand, $M_1, q_{NN} \not\models_{ir} \langle\langle 1, 2\rangle\rangle \Box \neg \mathsf{danger}$

because the only way to avoid a successful attack right after the game begins is that one guard stays, and one moves to the other position. Since they use memoryless strategies, the staying guard must execute *st* forever. Moreover, the moving guard will not see that he has changed his position (as $q_{NN} \sim_1 q_{SN}$ and $q_{NN} \sim_2 q_{NS}$). Since they can only use uniform strategies, the moving guard must execute *mv* forever – but that means that the area can be successfully attacked in two steps from the start. Finally, memory matters: $M_1, q_{NN} \models_{\mathrm{iR}} \langle\langle 1,2\rangle\rangle\Box\neg\mathsf{danger}$. To see this, consider any strategy where one guard always stays, and the other one moves in the first moment, and stays from then on. We leave it up to the reader to check that the strategy succeeds from $\{q_{NN}, q_{SS}\}$, i.e., both states that the guards jointly consider possible at the beginning.

As a *logic*, we will understand the language together with the chosen semantic interpretation. For the logics used in this paper, we will use the notation $L_{xy} = (\mathcal{L}_L, \models_{\mathrm{xy}})$. For example, 1ATL$_{\mathrm{ir}}$ denotes the logic with the syntax defined by $\mathcal{L}_{\mathrm{1ATL}}$ and the semantics by $\models_{\mathrm{ir}}$.

3 A Different Concept of Coalitional Uniformity

The uniformity conditions presented in Section 2.3 are based on the assumption that each member of the coalition executes its part of the joint plan on its own. Thus, the execution of every next step is based on the agent's individual view of the situation. A number of papers redefine uniformity of coalitional strategies, using instead a *single* epistemic relation for the whole coalition [13, 11, 14, 10]. This amounts to assuming members of the coalition to establish their *joint view of the situation* at each step while executing the joint strategy. Thus, at each step they either fully share their individual knowledge, or aggregate their uncertainty. [13, 11, 14] take the first approach by defining coalitional uniformity on top of the distributed knowledge relation. In [10], the opposite stance is adopted, by assuming that members of a coalition must choose same actions in states that are connected by the common knowledge relation. The main motivation for the semantic variations was quest for a variant of ATL with imperfect information, perfect recall, and decidable model checking.[5] However, the research was conceptually interesting in its own right.

We formalize the intuitions from [13, 11, 14, 10] by changing the set of available strategies as follows.

Definition 2 (Coalition-uniform strategies). *A collective memoryless strategy f_A is* coalition-uniform *iff $q \sim_A q'$ implies $f_a(q) = f_a(q')$ for every $q, q' \in St$ and $a \in A$. Likewise, a collective perfect recall strategy f_A is* coalition-uniform *iff $h \approx_A h'$ implies $f_a(h) = f_a(h')$ for every $h, h' \in St^+$ and $a \in A$.*

Note that uniformity constraints are relevant only for the imperfect information case. Depending on the type of recall, we will denote the new variant of ATL

[5] It is well known that "standard" ATL with imperfect information and perfect recall makes verification undecidable even for very simple formulae and regular models [12].

by ATL^c_{ir} or ATL^c_{iR}, with "c" standing for *imperfect information with **c**oalition-uniform strategies*. Let $y \in \{r, R\}$. The interpretation of strategic modalities in ATL^c_{iy} is defined below:

- $M, q \models^c_{\text{iy}} \langle\langle A\rangle\rangle \bigcirc \varphi$ iff there exists a set of coalition-uniform y-strategies F_A, such that for all q' with $q \sim_A q'$ and all computations $\lambda \in out(q', F_A)$ we have $\lambda[1] \models^c_{\text{iy}} \varphi$.
- $M, q \models^c_{\text{iy}} \langle\langle A\rangle\rangle \Box \varphi$ iff there exists a set of coalition-uniform y-strategies F_A, such that for all q' with $q \sim_A q'$ and all computations $\lambda \in out(q', F_A)$ we have $\lambda[i] \models^c_{\text{iy}} \varphi$ for all i.
- $M, q \models^c_{\text{iy}} \langle\langle A\rangle\rangle \varphi_1 \mathcal{U} \varphi_2$ iff there exists a set of coalition-uniform y-strategies F_A such that $\forall q' \sim_A q$ and $\forall \lambda \in out(q', F_A)$, there exists a position $i \geq 0$ such that $\lambda[i] \models^c_{\text{iy}} \varphi_2$, and for all positions $0 \leq j < i$, we have $\lambda[j] \models^c_{\text{iy}} \varphi_1$.

In line with [13, 11, 14], we assume $\sim_A = \sim^D_A$ and $\approx_A = \approx^D_A$. That is, members of a coalition are able to freely communicate while executing the strategy. We believe, however, that our results carry over to the other notions of collective indistinguishability.

Example 4. Now, we have that $M_1, q_{NN} \models^c_{\text{ir}} \langle\langle 1, 2\rangle\rangle \Box \neg \mathsf{danger}$. A successful strategy makes one guard do *st* in every state, and the other guard move in $\{q_{NN}, q_{SS}\}$ and stay elsewhere.

4 Translating Coalition-Uniform ATL to Single-Agent ATL

In this section we present a truth-preserving translation from ATL^c_{iy} to 1ATL_{iy}.

4.1 Reconstruction of Models

We first propose a reconstruction of ICGS's that replaces relevant coalitions by single agents. The idea is as follows: for every coalition A occurring in a given formula φ, we remove the agents in A from Σ, and instead add a new agent a_A. The actions of a_A are combinations of actions from agents in A. Thus, the new set of agents will consist of new agents representing coalitions from φ, plus those agents that did not appear in any coalition. Now, it can be the case that some "old" agents have become part of several different "coalitional" agents a_A. If their choices agree across the new coalitional actions then the transition specified in the original model is executed. Otherwise, the system proceeds to a new "conflict" state $q_\perp$ labeled with a new atomic proposition null.

Definition 3 (Model Translation). *We define a function T which given an ICGS $M = \langle \Sigma, \Pi, Q, C, d, \delta, \sim, \pi\rangle$ and an ATL^c_{iy} formula φ with coalitions $A_1, \ldots, A_n$ of (where $A_1, \ldots, A_n \subseteq \Sigma$), translates them into a concurrent game structure $T(M, \varphi) = M' = \langle \Sigma', \Pi', Q', C', d', \delta', \sim', \pi'\rangle$.*

$\Sigma' = \{a_{A_1}, \ldots, a_{A_n}\} \cup \{\Sigma \setminus \bigcup_n A_n\} \cup \{a_d\}$ *is the new set of agents, which has new agents $\{a_{A_1}, \ldots, a_{A_n}\}$ that correspond to coalitions occurring in φ, and all* *the old agents except for those that belonged to coalitions in φ. There is also an extra agent in Σ' which we denote as 'a_d' ("dummy"). For the sake of brevity we will refer to 'old' agents (those that do not belong to coalitions occurring in φ) as $\{a_1, \ldots, a_m\}$. Also, in order to be able to refer to former members of coalitions $A_1, \ldots, A_n$, we adopt the following notation:*

$$\begin{aligned} A_1 &= (a_1^1, a_2^1, \ldots, a_{l_1}^1) \\ A_2 &= (a_1^2, a_2^2, \ldots, a_{l_2}^2) \\ &\vdots \\ A_n &= (a_1^n, a_2^n, \ldots, a_{l_n}^n) \end{aligned}$$

And we also say that there are k agents in the original structure M.

We introduce one new propositional symbol 'null' and one new state $q_\perp$:

$$\begin{aligned} \Pi' &= \Pi \cup \{null\} \\ Q' &= Q \cup \{q_\perp\} \end{aligned}$$

We now define the set of choices C' and a function of enabled choices d' simultaneously. We say that:

$$d'_a(q') = \begin{cases} d_a(q) & \text{if } a \in \{a_1, \ldots, a_m\} \text{ and } q' \neq q_\perp, \\ \prod_{b \in A_j} d_b(q) & \text{if } a = a_{A_j} \text{ and } q' \neq q_\perp, \\ \{empty\} & \text{if } a = a_d \text{ or if } q' = q_\perp, \end{cases}$$

where $q' \in Q'$, $q \in Q$ and $a \in \Sigma'$. We say the set C' is simply an image of the function d', and we refer to members of C' as c'.

The new transition function, δ', handles non-empty intersections of coalitions that lead to (potentially) conflicting choices. Whenever two (or more) singleton coalitions have enabled choices that would lead to different states, it produces transitions to a special conflict state $q_\perp$:

$$\delta'(q', (c'_{a_1^1}, \ldots, c'_{a_{l_1}^1}), \ldots, (c'_{a_1^n}, \ldots, c'_{a_{l_n}^n}),$$
$$c'_d, c'_1, \ldots, c'_m) = \begin{cases} \delta(q, x_1, \ldots, x_k) & \text{when } A_1 \cap \ldots \cap A_n = \emptyset \text{ or} \\ & \quad \forall_{i,j \in \{1,\ldots,n\}} \forall_{r \in \{1,\ldots,l_i\}} \forall_{s \in \{1,\ldots,l_j\}} \\ & \quad\quad a_r^i = a_s^j \Rightarrow c'_{a_r^i} = c'_{a_s^j} \\ q_\perp & \text{otherwise.} \end{cases}$$

where $x_i = c'_j$ if $a_i = a_j \in \{a_1, \ldots, a_m\}$ and $c'_{a_o^j}$ if $a_i = a_o^j \in A_j$.

The indistinguishability relation remains the same for old agents, and for new agents it becomes the intersection of relations for members of old coalitions:

$$\sim'_a = \begin{cases} \sim_a \cup \{q_\perp, q_\perp\} & \text{if } a \in \{a_1, \ldots, a_n\} \\ \bigcap_{b \in A_j} \sim_b & \text{if } a = a_{A_j} \\ \{q_\perp, q_\perp\} & \text{if } a = a_d \end{cases}$$

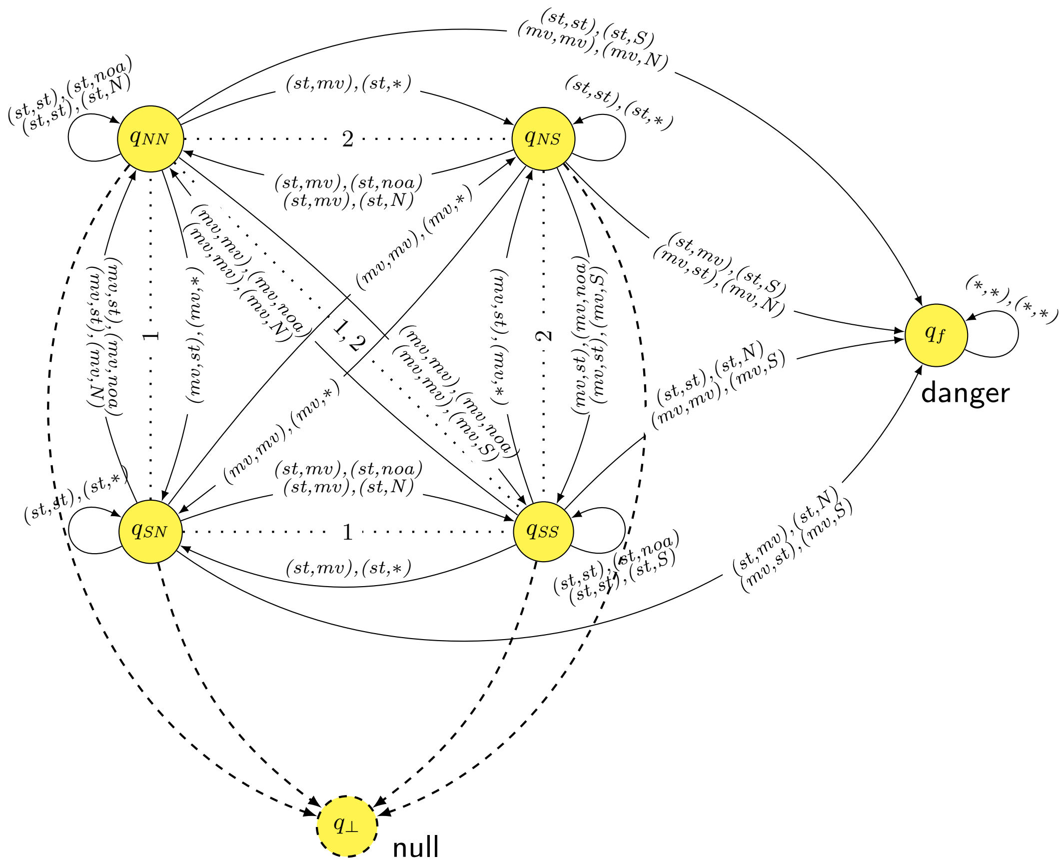

Fig. 2. Translation of model M_1 for formula $\langle\langle 1,2\rangle\rangle \Diamond \langle\langle 1,a\rangle\rangle \bigcirc \mathsf{danger}$. Dashed arrows depict transitions to the "conflict" state $q_\perp$

Finally, we label the special state $q_\perp$ with a new propositional symbol null: $\pi'(q) = \{null\}$ for $q = q_\perp$ and $\pi(q)$ otherwise.

Example 5. The translation of the coordinated defense model M_1 for formula $\langle\langle 1,2\rangle\rangle \Diamond \langle\langle 1,a\rangle\rangle \bigcirc \mathsf{danger}$ is presented in Figure 2.

4.2 Translation of Formulas

The formula translation is straightforward. We substitute each coalition A with agent a_A, and insert the proposition null so that the opponents can only lose by enforcing a conflict.

Definition 4 (Formula Translation). *We define the function* $t : \mathcal{L}_{\text{ATL}} \to \mathcal{L}_{\text{1ATL}}$ *which translates an* ATL *formula over* Σ *to a single-agent* ATL *formula over* Σ' *inductively in the following way:*

$$\begin{aligned}
t(p) &= p \\
t(\neg\varphi) &= \neg t(\varphi) \\
t(\varphi \wedge \psi) &= t(\varphi) \wedge t(\psi) \\
t(\langle\langle\emptyset\rangle\rangle \bigcirc \varphi) &= \langle\langle a_d\rangle\rangle \bigcirc (null \vee t(\varphi)) \\
t(\langle\langle\emptyset\rangle\rangle \Box \varphi) &= \langle\langle a_d\rangle\rangle \Box (null \vee t(\varphi)) \\
t(\langle\langle\emptyset\rangle\rangle \varphi \mathcal{U} \psi) &= \langle\langle a_d\rangle\rangle (null \vee t(\varphi)) \mathcal{U} (null \vee t(\psi)) \\
t(\langle\langle A\rangle\rangle \bigcirc \varphi) &= \langle\langle a_A\rangle\rangle \bigcirc (null \vee t(\varphi)) \\
t(\langle\langle A\rangle\rangle \Box \varphi) &= \langle\langle a_A\rangle\rangle \Box (null \vee t(\varphi)) \\
t(\langle\langle A\rangle\rangle \varphi \mathcal{U} \psi) &= \langle\langle a_A\rangle\rangle (null \vee t(\varphi)) \mathcal{U} (null \vee t(\psi))
\end{aligned}$$

where $p \in \Pi$*,* $null \in \Pi'$*,* $A \subseteq \Sigma$*, and* $a_A, a_d \in \Sigma'$*.*

Example 6. According to the translation, our formula $\langle\langle 1,2\rangle\rangle \Diamond \langle\langle 1,a\rangle\rangle \bigcirc \mathsf{danger}$ becomes now $\langle\langle a_{\{1,2\}}\rangle\rangle \Diamond \big(null \vee \langle\langle a_{\{1,a\}}\rangle\rangle \bigcirc (null \vee \mathsf{danger})\big)$.

4.3 Main Result: The Embedding Is Truth-Preserving

In order to prove correctness of our translation we need some additional definitions and lemmas:

Definition 5 (Complexity of formulas). *The complexity* $c : \mathcal{L}_{\text{ATL}^c_{\text{xy}}} \to \mathbb{N}$ *is defined inductively as follows:*

$$\begin{aligned}
c(p) &= 1 \\
c(\neg\varphi) &= 1 + c(\varphi) \\
c(\varphi \wedge \psi) &= 1 + \max(c(\varphi), c(\psi)) \\
c(\langle\langle A\rangle\rangle \bigcirc \varphi) = c(\langle\langle\emptyset\rangle\rangle \bigcirc \varphi) &= 1 + c(\varphi) \\
c(\langle\langle A\rangle\rangle \Box \varphi) = c(\langle\langle\emptyset\rangle\rangle \Box \varphi) &= 1 + c(\varphi) \\
c(\langle\langle A\rangle\rangle \varphi \mathcal{U} \psi) = c(\langle\langle\emptyset\rangle\rangle \varphi \mathcal{U} \psi) &= 1 + c(\varphi) + c(\psi)
\end{aligned}$$

Lemma 1. $T(M, \neg\varphi)$ *is equivalent to* $T(M, \varphi)$*.*

Proof. T takes a structure and a formula as its arguments, but the only part of the formula taken into consideration is the coalition predicate. Since negation does not affect it, we say that the above expressions are equivalent. □

We can now state the main result in this paper.

Theorem 1. *Given an* ICGS M*, a state* $q \in M$ *and an* ATL^c_{iy} *formula* φ*, the following equivalence holds:*

$$M, q \models \varphi \iff T(M, \varphi), q \models t(\varphi)$$

Proof. Let φ be a formula. The proof is by induction on $c(\psi')$ for all subformulas ψ' of φ.

Base case: If $M, q \models p$ then $T(M,\varphi), q \models t(p)$, because $t(p)$ is p, and for every $q \in Q$, $\pi'(q) = \pi(q)$. The same argument applies in the other direction, so if $T(M,\varphi), q \models t(p)$ then $M, q \models p$.

Induction hypothesis: For all subformulas ψ of φ such that $c(\psi) < c(\psi')$, and all $q \in Q$: $M, q \models \psi \iff T(M,\varphi), q \models t(\psi)$.

Case for $\psi' = \neg\psi$: Follows directly from the induction hypothesis and Lemma 1.

Case for $\psi' = \psi_1 \wedge \psi_2$: If ψ' is a conjunction, it is trivially true that whenever $M, q \models \psi_1 \wedge \psi_2$, then $T(M,\varphi), q \models t(\psi_1 \wedge \psi_2)$. The other direction follows as well, since $t(\psi_1 \wedge \psi_2)$ translates into $t(\psi_1) \wedge t(\psi_2)$.

Case for $\psi' = \langle\langle A\rangle\rangle \bigcirc \psi$:

(⇒) We want to show that:

$$M, q \models \langle\langle A\rangle\rangle \bigcirc \psi \Rightarrow T(M,\varphi), q \models \langle\langle a_A\rangle\rangle \bigcirc (\mathit{null} \vee t(\psi))$$

Assume that $\exists F_A \forall_{q_1 \sim_A q} \forall \lambda \in \mathit{out}(q_1, F_A), M, \lambda[1] \models \psi$, where $F_A = \{f_a : a \in A\}$ is a coalition-uniform set of strategies. We must show that $\exists f_{a_A} \forall_{q_1' \sim'_{a_A} q} \forall \lambda' \in \mathit{out}(q_1', f_{a_A}), T(M,\varphi), \lambda'[1] \models (\mathit{null} \vee t(\psi))$. We define the strategy f_{a_A} for agent a_A in $T(M,\varphi)$ as follows: $f_{a_A}(q') = \prod_{a \in A} f_a(q')$ when $q' \in Q$, and $f_{a_A}(q') = \mathit{empty}$ when $q' = q_\perp$. Let $q_1' \sim'_{a_A} q$ and $\lambda' \in \mathit{out}(q_1', f_{a_A})$. We must show that $T(M,\varphi), \lambda'[1] \models (\mathit{null} \vee t(\psi))$. From Definitions 2 and 3 we have that $q_1' \sim'_{a_A} q$ implies that $q_1' \sim_A q$. That $\lambda' \in \mathit{out}(q_1', f_{a_A})$ means that there is a choice $c_a' \in d_a'(q_1')$ for each agent $a \in \Sigma'$ such that $c'_{a_A} = f_{a_A}(q_1')$. We now define some notation: for any $h \in \{1, \ldots, n\}$, we have that $c'_{a_{A_h}} = (c_1^h, \ldots, c_{l_h}^h) \in d'_{a_{A_h}}(q_1') = \prod_{a \in A_h} d_a(q_1')$, so for every agent $a \in \Sigma$ such that $a \in A_h = \{a_1^h, \ldots, a_{l_h}^h\}$, let $i_a^h \in \{1, \ldots, l_h\}$ be such that $a = a^h_{i_a^h}$ (a is agent number i_a^h in the enumeration of A_h). We thus have that $c^h_{i_a^h} \in d(q_1')$ is the choice of agent a in the choice $c'_{a_{A_h}}$ of the coalition A_h. We now consider two cases.

The first case is that there exist A_h and A_l such that $A_h \cap A_l \neq \emptyset$ but $c^h_{i_a^h} \neq c^l_{i_a^l}$ (the choice made by a in the two coalitions differ). In this case $\delta'(q_1', \mathbf{c}) = q_\perp$, and thus $T(M,\varphi), \lambda'[1] \models \mathit{null}$ and we are done.

Assume, then, the second case, that whenever there are one or more coalitions with a as a member, they all agree on the choice for agent a, i.e., $c^h_{i_a^h} = c^l_{i_a^l}$ whenever $a \in A_h \cap A_l$. We now define a choice $c_a \in d_a(q_1')$ for each agent $a \in \Sigma$ in the original model M, as follows. When $a \notin A_h$, for all h, let $c_a = c_a'$; $c_a \in d_a(q_1')$ because $d_a'(q_1') = d_a(q_1')$. When $a \in A_h$ for some h, let $c_a = c^h_{i_a^h}$ for some h such that $a \in A_h$ (this is well-defined by the assumption that all coalitions with a as a member agree on the choice of a). Let h be such that $A = A_h$ (since ψ' is a subformula of φ, A is one of the coalitions occurring in φ). Let $a \in A$. We have that $c'_{a_A} = f_{a_A}(q_1') = \prod_{a \in A} c_a$ and $c_a = c^h_{i_a^h}$ by the definitions above, and thus $c_a = c^h_{i_a^h} = f_a(q_1')$. Since $c_a = f_a(q_1')$ for any $a \in A$, there is a $\lambda \in \mathit{out}(q_1', F_A)$

such that $\lambda[1] = \delta(q'_1, \mathbf{c})$, and we thus have that $M, \delta(q'_1, \mathbf{c}) \models \psi$.[6] By the induction hypothesis, $T(M, \varphi), \delta(q'_1, \mathbf{c}) \models t(\psi)$. By Definition 3 we have that $\delta'(q'_1, \mathbf{c}') = \delta(q'_1, \mathbf{c})$, and thus that $T(M, \varphi), \delta(q'_1, \mathbf{c}) \models t(\psi)$. By definition of $\mathbf{c}'$, $\delta'(q'_1, \mathbf{c}') = \lambda'[1]$. Thus, $T(M, \varphi), \lambda'[1] \models t(\psi)$, and we are done.

(⇐) We want to show that:

$$T(M, \varphi), q \models \langle\langle a_A \rangle\rangle \bigcirc (null \vee t(\psi)) \Rightarrow M, q \models \langle\langle A \rangle\rangle \bigcirc \psi$$

Assume that $\exists f_{a_A} \forall_{q_1 \sim'_{a_A} q} \forall \lambda' \in out(q_1, f_{a_A}), T(M, \varphi), \lambda'[1] \models (null \vee t(\psi))$. We must show that $\exists F_A \forall_{q_1 \sim_A q} \forall \lambda \in out(q_1, F_A), M, \lambda[1] \models \psi$.

We define a coalition-uniform set of strategies $F_A = \{f_a : a \in A\}$ for coalition $A = \{a_1, \ldots, a_r\}$ as follows: for every $a = a_j \in A$ and any $q' \in Q$, $f_a(q') = c_j$, where $(c_1, \ldots, c_r) = f_{a_A}(q)$. For $q' = q_\perp$ and $a \in A$, $f_a(q') = empty$. From Definition 3 it is easy to see that F_A is a (collective) strategy in M (i.e., $f_a(q') \in d_a(q')$ for each $a \in A$) and from uniformity of f_{a_A} and Definition 3 it follows that F_A is coalition-uniform ($q' \sim_A q'' \Rightarrow f_a(q') = f_a(q'')$ for each $a \in A$).

Let $q_1 \sim_A q$. We know that $q \neq q_\perp$, because $q_\perp \notin Q$ and we also know that $q_1 \neq q_\perp$, because $q_\perp$ is indistinguishable only from itself. Hence, from Definitions 2 and 3 we get that $q_1 \sim_A q$ implies $q_1 \sim'_{a_A} q$.

Let $\lambda \in out(q_1, F_A)$. It is easy to see that also $\lambda \in out(q_1, f_{a_A})$: $T(M, \varphi)$ includes all the states of M; all the strategies F_A and f_{a_A} "do the same thing" in those states; the other agents have the same actions available in those states. Thus, $T(M, \varphi), \lambda[1] \models null \vee t(\psi)$. But it cannot be that $T(M, \varphi), \lambda[1] \models null$, because $null$ is only satisfied in $q_\perp$ and $q_\perp$ is not a state in λ (since λ is a computation in M). So, $T(M, \varphi), \lambda[1] \models t(\psi)$, and by the induction hypothesis, $M, \lambda[1] \models \psi$. Thus, $M, q \models \langle\langle A \rangle\rangle \bigcirc \psi$.

Case for $\psi' = \langle\langle A \rangle\rangle \psi_1 \mathcal{U} \psi_2$:

(⇒) The proof of this case proceeds in a similar way to the previous case. We want to show that:

$$M, q \models \langle\langle A \rangle\rangle \psi_1 \mathcal{U} \psi_2 \Rightarrow T(M, \varphi), q \models \langle\langle a_A \rangle\rangle (null \vee t(\psi_1)) \mathcal{U} (null \vee t(\psi_2))$$

Assume that $\exists F_A$ such that $\forall_{q_1 \sim_A q} \forall \lambda \in out(q_1, F_A)$, there is a position $i > 0$ in λ, such that $M, \lambda[i] \models \psi_2$ and for all positions $0 \leq j < i$, $M, \lambda[j] \models \psi_1$, where $F_A = \{f_a : a \in A\}$ is a coalition-uniform set of strategies. We must show that $\exists f_{a_A} \forall_{q_1 \sim'_{a_A} q} \forall \lambda' \in out(q'_1, f_{a_A})$, there is a position $i' > 0$ in λ', such that $T(M, \varphi), \lambda'[i'] \models (null \vee t(\psi_2))$ and for all positions $0 \leq j' < i'$, $T(M, \varphi), \lambda'[j'] \models (null \vee t(\psi_1))$. We define the strategy f_{a_A} for agent a_A in $T(M, \varphi)$ like before: $f_{a_A}(q') = \prod_{a \in A} f_a(q')$ when $q' \in Q$, and $f_{a_A}(q') = empty$ when $q' = q_\perp$. Let $q_1 \sim'_{a_A} q$ and $\lambda' \in out(q_1, f_{a_A})$.

It is now easy to see, in the same way as in the $\bigcirc$-case, that there exists an M-computation $\lambda \in out(q_1, F_A)$ such that either (i) $\lambda[j] = \lambda'[j]$ for all

[6] Throughout the rest of the proof we use the following notation: $\mathbf{c}$ is the action profile where the choice of agent $a \in \Sigma$ is c_a and $\mathbf{c}'$ is the action profile where the choice of agent $a \in \Sigma'$ is c'_a.

$j \geq 0$, or (ii) there exists a $k \geq 0$ such that $\lambda[j] = \lambda'[j]$ for all $0 \leq j < k$ and $\lambda'[j] = q_\perp$ for all $j \geq k$. In case (i) we are done by the induction hypothesis. We argue that we are also done in case (ii). If $k > i$ where i is such that $M, \lambda[i] \models \psi_2$, we are done by the induction hypothesis like in case (i). If $k \leq i$ we are also done: we have that $M, \lambda[j] \models \psi_1$ for all $j \leq k$; by the induction hypothesis $T(M, \varphi), \lambda[j] \models t(\psi_1)$ for all $j \leq k$ and thus $T(M, \varphi), \lambda'[j] \models t(\psi_1)$ for all $j \leq k$; and $T(M, \varphi), \lambda'[k] \models null$.

(⇐) The proof in this direction is exactly like the proof in the same direction for the $\bigcirc$-case.

Case for $\psi' = \langle\langle A \rangle\rangle \Box \psi$: analogous to the $\mathcal{U}$-case. □

5 Conclusions

In this paper, we look closer at the issue of executable strategies for coalitions acting under imperfect or incomplete information. Based on ideas from existing literature, we propose the "coalition-uniform" semantics for Alternating-time Temporal Logic where uniformity of coalitional strategies is based on the *distributed knowledge* relation for the coalition. We also show that ATL with the new semantics can be embedded in the syntactic restriction of the logic that talks only about abilities of individual agents. This is done through a translation of models and formulae that preserves the truth of formulae in the context of a given model. We take it as a formal counterpart of our argument that coalitions whose members can fully coordinate their actions and share their knowledge should be seen as *de facto* single compound agents. We also note that the translation can be used to implement model checking of coalition-uniform ATL with verification tools for more standard variants of the logics, such as MCMAS [19] and SMC [20].

Acknowledgements. Piotr Kaźmierczak's research was supported by the Research Council of Norway project 194521 (FORMGRID). Wojciech Jamroga acknowledges the support of the National Research Fund (FNR) Luxembourg under the project GALOT (INTER/DFG/12/06), as well as the support of the 7th Framework Programme of the European Union under the Marie Curie IEF project ReVINK (PIEF-GA-2012-626398).

References

1. T. Ågotnes. Action and knowledge in alternating-time temporal logic. *Synthese*, 149(2):377–409, 2006. Section on Knowledge, Rationality and Action.
2. T. Ågotnes, V. Goranko, and W. Jamroga. Alternating-time temporal logics with irrevocable strategies. In D. Samet, editor, *Proceedings of TARK XI*, pages 15–24, 2007.
3. N. Alechina, B. Logan, H. Nguyen, and A. Rakib. Resource-bounded alternating-time temporal logic. In *Proceedings of AAMAS*, pages 481–488, 2010.

4. R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 100–109. IEEE Computer Society Press, 1997.
5. R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. *Journal of the ACM*, 49:672–713, 2002.
6. T. Brihaye, A. D. C. Lopes, F. Laroussinie, and N. Markey. ATL with strategy contexts and bounded memory. In *Proceedings of LFCS*, volume 5407 of *Lecture Notes in Computer Science*, pages 92–106. Springer, 2009.
7. N. Bulling and J. Dix. Modelling and verifying coalitions using argumentation and ATL. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, 14(46):45–73, 2010.
8. N. Bulling and B. Farwer. On the (un-)decidability of model checking resource-bounded agents. In *Proceedings of ECAI*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, pages 567–572. IOS Press, 2010.
9. K. Chatterjee, T. A. Henzinger, and N. Piterman. Strategy logic. In *Proceedings of CONCUR*, pages 59–73, 2007.
10. R. Diaconu and C. Dima. Model-checking alternating-time temporal logic with strategies based on common knowledge is undecidable. *Applied Artificial Intelligence*, 26(4):331–348, 2012.
11. C. Dima, C. Enea, and D. Guelev. Model-checking an alternating-time temporal logic with knowledge, imperfect information, perfect recall and communicating coalitions. In *Proceedings of GANDALF*, pages 103–117, 2010.
12. C. Dima and F. Tiplea. Model-checking atl under imperfect information and perfect recall semantics is undecidable. *CoRR*, abs/1102.4225, 2011.
13. D. Guelev and C. Dima. Model-checking strategic ability and knowledge of the past of communicating coalitions. In *Proceedings of DALT*, pages 75–90, 2008.
14. D. Guelev, C. Dima, and C. Enea. An alternating-time temporal logic with knowledge, perfect recall and past: axiomatisation and model-checking. *Journal of Applied Non-Classical Logics*, 21(1):93–131, 2011.
15. W. Jamroga. A temporal logic for stochastic multi-agent systems. In *Proceedings of PRIMA'08*, volume 5357 of *LNCS*, pages 239–250, 2008.
16. W. Jamroga and T. Ågotnes. Constructive knowledge: What agents can achieve under incomplete information. *Journal of Applied Non-Classical Logics*, 17(4):423–475, 2007.
17. W. Jamroga and A. Murano. On module checking and strategies. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems AAMAS 2014*, pages 701–708, 2014.
18. W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63(2–3):185–219, 2004.
19. A. Lomuscio, H. Qu, and F. Raimondi. MCMAS : A model checker for the verification multi-agent systems. In *Proceedings of CAV*, volume 5643 of *Lecture Notes in Computer Science*, pages 682–688, 2009.
20. J. Pilecki, M. Bednarczyk, and W. Jamroga. Synthesis and verification of uniform strategies for multi-agent systems. In *Proceedings of the 15th Workshop on Computational Logic in Multi-Agent Systems CLIMA XV*, volume 8624 of *LNCS*, pages 166–182. Springer, 2014.
21. H. Schnoor. Strategic planning for probabilistic games with incomplete information. In *Proceedings of AAMAS'10*, pages 1057–1064, 2010.
22. P. Y. Schobbens. Alternating-time logic with imperfect recall. *Electronic Notes in Theoretical Computer Science*, 85(2):82–93, 2004.

23. W. van der Hoek, W. Jamroga, and M. Wooldridge. A logic for strategic reasoning. In *Proceedings of AAMAS'05*, pages 157–164, 2005.
24. W. van der Hoek and M. Wooldridge. Cooperation, knowledge and time: Alternating-time Temporal Epistemic Logic and its applications. *Studia Logica*, 75(1):125–157, 2003.
25. S. van Otterloo and G. Jonker. On Epistemic Temporal Strategic Logic. *Electronic Notes in Theoretical Computer Science*, 126:77–92, 2004. Proceedings of LCMAS'04.
26. S. van Otterloo, W. van der Hoek, and M. Wooldridge. Preferences in game logics. In *Proceedings of AAMAS-04*, pages 152–159, 2004.
27. D. Walther, W. van der Hoek, and M. Wooldridge. Alternating-time temporal logic with explicit strategies. In D. Samet, editor, *Proceedings TARK XI*, pages 269–278. Presses Universitaires de Louvain, 2007.
28. M. Wooldridge. *An Introduction to Multi Agent Systems*. John Wiley & Sons, 2002.