Complexity of a theory of collective attitudes in teamwork

Marcin Dziubiński
Institute of Informatics,
Warsaw University,
Banacha 2,
02-097 Warsaw, Poland
amosild@mimuw.edu.pl

Rineke Verbrugge
Institute of Artificial Intelligence,
University of Groningen,
Grote Kruisstraat 2/1,
9712 TS Groningen,
The Netherlands
rineke@ai.rug.nl

Barbara Dunin-Kęplicz
Institute of Informatics,
Warsaw University,
Banacha 2,
02-097 Warsaw, Poland
keplicz@mimuw.edu.pl
Institute of Computer Science,
Polish Academy of Sciences,
Ordona 21,
01-237 Warsaw, Poland

Abstract

Our previous research presents a methodology of cooperative problem solving for BDI systems, based on a complete formal theory. This covers both a static part, defining individual, bilateral and collective agent attitudes, and a dynamic part, describing system reconfiguration in a dynamic, unpredictable environment. In this paper, we investigate the complexity of the satisfiability problem of the static part of our theory, focusing on individual and collective attitudes up to collective intention. Our logics for teamwork are squarely multi-modal, in the sense that different operators are combined and may interfere. One might expect that such a combination is much more complex than the basic multi-agent logic with one operator, but in fact we show that the individual part of our theory of teamwork is only PSPACE-complete. The full system, modeling a subtle interplay between individual and group attitudes, turns out to be EXPTIME-complete, and remains so even if propositional dynamic logic is added to it.

1. Introduction

In this paper, we investigate the complexity of some important teamwork logics constructed in our previous papers. Although the results and methods of this paper may be applied to a complexity analysis of many multi-modal logics combining different but interrelated agent attitudes, the jumping point of the paper is our own theory of teamwork as presented in [4].

1.1. A formal theory of mental states in teamwork

When constructing belief-desire-intention (BDI) systems, the first research question has been to create a model of an agent as an *individual, autonomous* entity. A more recent goal has been to organize agents' cooperation in a way allowing the achievement of their, possibly complex, common goal, while preserving, at least partly, the autonomy of particular agents involved. The BDI-model naturally comprises such individual notions like *beliefs*, referring to the agent's informational attitudes, as well as *goals* and *intentions*, dealing with its motivational stance. However in teamwork, when a team of agents needs to work together in a planned and coherent way, these are not enough: the group as a whole needs to present a common collective attitude over and above individual attitudes of team members. Without this, a sensible organization of cooperation seems to be impossible: the existence of collective (or joint) motivational attitudes is a necessary condition for a group of agents to become a cooperative team [4, 6]. Thus, in the context of teamwork agents' attitudes are considered on the individual, social (i.e. bilateral) and collective level. A theory of informational attitudes on the group level has been formalized in terms of epistemic logic [7]. As regards motivational attitudes, the situation is much more complex: a conceptually coherent theory was vitally needed in the MAS literature, as the notions on the bilateral and collective level cannot be viewed as a straightforward extension or a sort of sum of individual notions. In order to define them, many subtle aspects that are hard to formalize need to be introduced. A departure point to construct a *static, descriptive* theory of collective motivational attitudes is formed by individual goals, beliefs and intentions of cooperating agents. Research on teamwork should address the question what it means for a group of agents to have a *collective intention*, and then a *collective commitment* to achieve a common goal. In our approach, the fundamental role of collective intention is to consolidate a group as a cooperating team, while collective commitment leads to team action, i.e., to coordinated realization of individual actions by the agents that have committed to do them according to the team plan. Both notions are constructed in a way that allows us to fully express the potential of strictly cooperative teams [4, 6]. In this paper we will focus on the theory of teamwork built up from the above-mentioned individual and collective attitudes up to collective intentions, where our goal is to give an exact complexity class of the full logic and its most important subsystem for individual attitudes. When modeling collective intention, agents' awareness about the overall situation needs to be taken into account. In a theory of multiagent systems, the essential notion of awareness is understood as a reduction of the general sense of this notion to the state of an agent's beliefs about itself, about other agents and about the state of the environment. Under this assumption, various epistemic logics and various notions of group information (from distributed belief to common knowledge) are adequate to formalize agents' awareness [7, 6]. In the presented theory, group awareness is usually expressed in terms of the (rather strong) notion of *common belief*, but one can also consider weaker forms depending on the circumstances in question. In the complexity analysis, we do *not* take into account the cognitive and other processes necessary for establishing attitudes that appear in the definitions found in section 2. We are just interested in showing how complex it is to check *satisfiability* and validity of the formulas with respect to our theory. Let us turn to a short reminder about the decidability and complexity of such important questions about logical theories.

1.2. Complexity of logical tasks

In this paper we investigate the complexity of two modal logics for multiagent systems. In particular, we examine the complexity of their *satisfiability problem*: given a formula φ, how much time and space (in terms of the length of φ) are needed to compute whether φ is satisfiable, i.e. whether there is a suitable Kripke model $\mathcal{M}$ (from the class of structures corresponding to the logic) and a world s in it, such that $\mathcal{M}, s \models \varphi$? From this, the complexity of the *validity problem* (truth in all worlds in all suitable Kripke models) follows immediately, because φ is valid if and only if $\neg\varphi$ is not satisfiable. *Model checking*, i.e. evaluating truth of a given formula in a given world and model ($\mathcal{M}, s \models \varphi$) is the most important related problem, and is easily seen to be less complex than both satisfiability and validity. We do not investigate the complexity of model checking here, but see [12] for such an analysis of some MAS logics; in any case, various methods have already been developed that can perform model checking in a reasonable time, as long as the considered models are not too large.

Unfortunately, even satisfiability for propositional logic is NP-complete. Thus, if P $\neq$ NP, then modal logics interesting for MAS, all containing propositional logic as a subsystem, do not have efficiently solvable satisfiability problems. Even though a single efficient algorithm performing well on all inputs is not possible, it is still important to discover in which complexity class a given logical theory falls. In our work we take the point of view of the system developer who wants to reason about, specify and verify a MAS to be constructed. It turns out that for many of the interesting formulas appearing in such human reasoning, satisfiability tends to be easier to compute than suggested by worst-case labels like "PSPACE-complete" and "EXPTIME-complete" [9]. It would be helpful to develop automated, efficient tools to support the system developer with some reasoning tasks, and in section 5 we will come back to methods that can simplify satisfiability problems for MAS logics in an application-dependent way.

Of many single-agent modal logics with one modality, the complexity has long been known. An overview is given in [9], which extends these results to multi-agent logics, though still containing only a single modality (either knowledge or belief). For us, the following results are relevant. The satisfiability problems for the systems $\mathrm{S5}_1$ and $\mathrm{KD45}_1$, modeling knowledge and belief of one agent, are NP-complete. Thus, surprisingly, they are no more complex than propositional logic. The complexity is increased to PSPACE if these systems are extended to more than one agent. PSPACE is also the complexity class of satisfiability for many other modal logics, for both the single and the multiagent case; examples are the basic system K_n (that we adopt for goals) and the system KD_n (for intentions, see section 2). As soon as a notion of seemingly infinite character such as common knowledge or common belief (everybody believes and everybody believes that everybody believes and ...) is modeled, the complexity of the satisfiability problem jumps to EXPTIME. Intuitively, trying to find a satisfying model for a formula containing a common belief operator by the tableau method, one may need to look exponentially deep in the tableau tree to find it, while for simpler modal logics like K_n, a depth-first search through a polynomially shallow tree suffices for all formulas.

When investigating the complexity of multi-modal logics, one might like to turn to general results on the transfer of the complexity of satisfiability problems from single logics to their combinations: isn't a combination of a few PSPACE-complete logics, with some simple interdependency axioms, automatically PSPACE-complete again? However, it turns out that the positive general results that do exist apply mainly to minimal combinations, without added interdependencies, of two NP-complete systems, each with a single modality. Even more dangerously, there are some very negative results on the transfer of complexity to combined systems. Thus, there are two "very decidable" logics whose combination, even without any interrelation axioms, is undecidable[1]. This goes to show that one needs to be very careful with any assumptions about generalizations of complexity results to combined systems.

Our logics for teamwork are squarely multi-modal, not only in the sense of modeling a multi-agent version of one modal operator, but also in the sense that different operators are combined and may interfere. One might expect that such a combination is much more complex than the basic multiagent logic with one operator, but in fact we show that this is not the case: the "individual part" of our theory of teamwork is only PSPACE-complete. In order to prove this, the semantic properties relating to the interdependency axioms must be carefully translated to conditions on the multimodal tableau with which satisfiability is tested. Of course the really challenging question appears when informational and motivational group notions are added. We show that also for this expressive system, modeling a subtle interplay between individual and group attitudes, satisfiability is EXPTIME-complete, as for the system only modeling common belief. As a bonus, it turns out that even adding dynamic logic (which is relevant for our study of the evolution of motivational attitudes in changing environments [5]) does not increase complexity beyond EXPTIME.

2. Logical background

As mentioned before, we propose the use of multi-modal logics to formalize agents' informational and motivational attitudes as well as actions they perform. In the present paper we only present axioms relating attitudes of agents with respect to *propositions*, not actions. A proposition reflects a particular state of affairs.

2.1. The language

Formulas are defined with respect to a fixed finite set of agents. The basis of the inductive definition is given in the following definition.

Definition 1 (Language) The language is based on the following two sets:

- a numerable set $\mathcal{P}$ of *propositional symbols*;
- a finite set $\mathcal{A}$ of *agents*, denoted by numerals $1, 2, \ldots, n$.

Definition 2 (Formulas) We inductively define a set $\mathcal{L}$ of formulas as follows.

F1 each atomic proposition $p \in \mathcal{P}$ is a formula;
F2 if φ and ψ are formulas, then so are $\neg\varphi$ and $\varphi \wedge \psi$;
F4 if φ is a formula, $i \in \mathcal{A}$, and $G \subseteq \mathcal{A}$, then the following are formulas:
epistemic modalities $\mathrm{BEL}(i, \varphi)$, E-$\mathrm{BEL}_G(\varphi)$, C-$\mathrm{BEL}_G(\varphi)$;
motivational modalities $\mathrm{GOAL}(i, \varphi)$, $\mathrm{INT}(i, \varphi)$, E-$\mathrm{INT}_G(\varphi)$, M-$\mathrm{INT}_G(\varphi)$, C-$\mathrm{INT}_G(\varphi)$.

2.2. Semantics based on Kripke models

Each Kripke model for the language $\mathcal{L}$ consists of a set of worlds, a set of accessibility relations between worlds, and a valuation of the propositional atoms, as follows.

Definition 3 (Kripke model) A Kripke model is a tuple $\mathcal{M} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, Val)$, such that

1. W is a set of possible worlds, or states;
2. For all $i \in \mathcal{A}$, it holds that $B_i, G_i, I_i \subseteq W \times W$. They stand for the accessibility relations for each agent with respect to *beliefs*, *goals*, and *intentions*, respectively. For example, $(s, t) \in B_i$ means that t is an epistemic alternative for agent i in state s.
3. $Val : \mathcal{P} \times W \to \{0, 1\}$ is the function that assigns the truth values to atomic propositions in states.

A Kripke frame $\mathcal{F}$ is defined as a Kripke model, but without the valuation function. At this stage, it is possible to define the truth conditions pertaining to the language $\mathcal{L}$. The expression $\mathcal{M}, s \models \varphi$ is read as "formula φ is satisfied by world s in structure $\mathcal{M}$". Define world t to be G_B*-reachable* (respectively G_I*-reachable*) from world s if there is a path of length ≥ 1 in the Kripke model from s to t along accessibility arrows B_i (respectively I_i) that are associated with members i of G.

Definition 4 (Truth definition)

- $\mathcal{M}, s \models p$ iff $Val(p, s) = 1$;
- $\mathcal{M}, s \models \neg\varphi$ iff $\mathcal{M}, s \not\models \varphi$;
- $\mathcal{M}, s \models \varphi \wedge \psi$ iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \psi$;
- $\mathcal{M}, s \models \mathrm{BEL}(i, \varphi)$ iff $\mathcal{M}, t \models \varphi$ for all t such that sB_it;
- $\mathcal{M}, s \models \mathrm{GOAL}(i, \varphi)$ iff $\mathcal{M}, t \models \varphi$ for all t such that sG_it;
- $\mathcal{M}, s \models \mathrm{INT}(i, \varphi)$ $\mathcal{M}, t \models \varphi$ for all t such that sI_it;

---

1 For A, take a variant of dynamic logic with two atomic programs, both deterministic; take ; and $\cap$ as only operators. Satisfiability for A, like that of propositional dynamic logic itself, is in EXPTIME. For B, take the logic of the global operator A (Always), defined as follows: $M, w \models A\varphi$ iff for all $v \in W$, $M, v \models \varphi$. Satisfiability for B is in NP. In [3] (see also [2, Theorem 6.31]), it is shown that the minimal combination of A and B is not only *not* in EXPTIME, but even undecidable in any finite time.

- $\mathcal{M}, s \models$ E-BEL$_G(\varphi)$ iff for all $i \in G$, $\mathcal{M}, s \models \mathrm{BEL}(i, \varphi)$;
- $\mathcal{M}, s \models$ C-BEL$_G(\varphi)$ iff $\mathcal{M}, t \models \varphi$ for all t that are G_B-reachable from s.
- $\mathcal{M}, s \models$ E-INT$_G(\varphi)$ iff for all $i \in G$, $\mathcal{M}, s \models \mathrm{INT}(i, \varphi)$;
- $\mathcal{M}, s \models$ M-INT$_G(\varphi)$ iff $\mathcal{M}, t \models \varphi$ for all t that are G_I-reachable from s.

2.3. Axiom systems for individual and collective attitudes

Let us give a reminder of our logical theory BGI_n for individual attitudes and their interdependencies (see subsections 2.3.1, 2.3.2), followed by our additional axioms and rules for group attitudes (see subsection 2.3.3). These axioms and rules, together forming our teamwork theory $BGI_n^{C,M}$, are fully explained in [4]. All axiom systems introduced here are based on a finite set of n agents.

2.3.1. Axioms for individual attitudes For beliefs we take the well-known system KD45$_n$:

P1 All instantiations of propositional tautologies;

PR1 From φ and $\varphi \to \psi$, derive ψ; (Modus Ponens)

A2 $\mathrm{BEL}(i, \varphi) \wedge \mathrm{BEL}(i, \varphi \to \psi) \to \mathrm{BEL}(i, \psi)$ (Belief Distribution)

A4 $\mathrm{BEL}(i, \varphi) \to \mathrm{BEL}(i, \mathrm{BEL}(i, \varphi))$ (Positive Introspection)

A5 $\neg\mathrm{BEL}(i, \varphi) \to \mathrm{BEL}(i, \neg\mathrm{BEL}(i, \varphi))$ (Negative Introspection)

A6 $\neg\mathrm{BEL}(i, \bot)$ (Consistency)

R2 From φ infer $\mathrm{BEL}(i, \varphi)$ (Belief Generalization)

For goals, we take the system K$_n$ and for intentions the system KD$_n$. By **A2**$_D$, **R2**$_D$ and **A2**$_I$, **R2**$_I$ we denote axioms of goal and intention distribution and rules corresponding to goal and intention generalization respectively. By **A6**$_I$ we denote intention consistency.

2.3.2. Interdependencies between attitudes

A7$_{IB}$ $\mathrm{INT}(i, \varphi) \to \mathrm{BEL}(i, \mathrm{INT}(i, \varphi))$ (Positive Introspection for Intentions)

A8$_{IB}$ $\neg\mathrm{INT}(i, \varphi) \to \mathrm{BEL}(i, \neg\mathrm{INT}(i, \varphi))$ (Negative Introspection for Intentions)

A9$_{ID}$ $\mathrm{INT}(i, \varphi) \to \mathrm{GOAL}(i, \varphi)$ (Intention implies goal)

Similarly, **A7**$_{DB}$ and **A8**$_{DB}$ stand for positive and negative introspection for goals. By BGI_n we denote the axiom system consisting of all the above axioms and rules for individual beliefs, goals and intentions as well as their interdependencies.

2.3.3. Axioms for group attitudes

C1 E-BEL$_G(\varphi) \leftrightarrow \bigwedge_{i \in G} \mathrm{BEL}(i, \varphi)$.
C2 C-BEL$_G(\varphi) \leftrightarrow$ E-BEL$_G(\varphi \wedge$ C-BEL$_G(\varphi))$
RC1 From $\varphi \to$ E-BEL$_G(\psi \wedge \varphi)$ infer $\varphi \to$ C-BEL$_G(\psi)$ (Induction Rule)
M1 E-INT$_G(\varphi) \leftrightarrow \bigwedge_{i \in G} \mathrm{INT}(i, \varphi)$.
M2 M-INT$_G(\varphi) \leftrightarrow$ E-INT$_G(\varphi \wedge$ M-INT$_G(\varphi))$
M3 C-INT$_G(\varphi) \leftrightarrow$ M-INT$_G(\varphi) \wedge$ C-BEL$_G($M-INT$_G(\varphi))$
RM1 From $\varphi \to$ E-INT$_G(\psi \wedge \varphi)$ infer $\varphi \to$ M-INT$_G(\psi)$ (Induction Rule)

By $BGI_n^{C,M}$ we denote the union of BGI_n with the above axioms and rules for general and common beliefs and general, mutual and collective intentions.

2.4. Correspondences between axiom systems and semantics

Most of the axioms above, as far as they do not hold on all frames like **A2**, correspond to well-known structural properties on Kripke frames. Thus, the axiom **A4** holds in a Kripke frame $\mathcal{F}$ iff all B_i relations are transitive; **A5** holds iff all B_i relations are Euclidean; and **A6** holds iff all B_i relations are serial (for proofs of these correspondences and correspondence theory in general, see [14]). As for the interdependencies, the semantic property corresponding to **A7**$_{IB}$ is $\forall s, t, u((sB_it \wedge tI_iu) \to sI_iu)$, analogously for **A7**$_{GB}$. The property that corresponds to **A8**$_{IB}$ is $\forall s, t, u((sI_it \wedge sB_iu) \to uI_it)$, analogously for **A8**$_{GB}$. Finally, for **A9**$_{IG}$ the corresponding semantic property is $G_i \subseteq I_i$. For proofs of these correspondences, see [6].

3. Complexity of the system BGI_n

We will show that the satisfiability problem for BGI_n is PSPACE-complete. First we present an algorithm for deciding satisfiability of a BGI_n formula φ working in polynomial space, thus showing that the satisfiability problem is in PSPACE. The construction of the algorithm, and related results are based on the method presented in [9]. The method is centered around well known notions of a *propositional tableau*, a *fully expanded propositional tableau* (a set that along with any formula ψ contained in it, contains also all its subformulas either in positive or negated form each) and a *tableau* designed for a particular system of multimodal logic. Let us give adaptations of the most important definitions from [9] as a reminder:

Definition 5 (Propositional tableau) A *propositional tableau* is a set T of formulas such that:

1. if $\neg\neg\psi \in T$ then $\psi \in T$;
2. if $\varphi \wedge \psi \in T$ then both $\varphi, \psi \in T$;
3. if $\neg(\varphi \wedge \psi) \in T$ then either $\neg\varphi \in T$ or $\neg\psi \in T$;
4. there is no formula ψ such that ψ and $\neg\psi$ are in T.

A set of formuls T is *blatantly inconsistent* if for some formula ψ, both ψ and $\neg\psi$ are in T.

In a tableau for a modal logic, for a given formula φ, $\mathrm{Sub}(\varphi)$ denotes the set of all subformulas of φ and $\neg\mathrm{Sub}(\varphi) = \mathrm{Sub}(\varphi) \cup \{\neg\psi : \psi \in \mathrm{Sub}(\varphi)\}$.

Definition 6 (BGI_n tableau) A BGI_n *tableau* $\mathcal{T}$ is a tuple $\mathcal{T} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, L)$, where W is a set of states, B_i, G_i, I_i are binary relations on W and L is a *labeling function* associating with each state $w \in W$ a set $L(w)$ of formulas, such that $L(w)$ is a propositional tableau. Here follow the two conditions that every modal tableau for our language must satisfy (see [9]):

1. If $\mathrm{BEL}(i,\varphi) \in L(w)$ and $(w,v) \in B_i$, then $\varphi \in L(v)$; similarly for $\mathrm{GOAL}(i,\varphi)$ w.r.t. G_i and $\mathrm{INT}(i,\varphi)$ w.r.t. I_i.
2. If $\neg\mathrm{BEL}(i,\varphi \in L(w)$, then there exists a v with $(w,v) \in B_i$ and $\neg\varphi \in L(v)$; similarly for $\mathrm{GOAL}(i,\varphi)$ w.r.t. G_i and $\mathrm{INT}(i,\varphi)$ w.r.t. I_i.

Furthermore, a BGI_n tableau must satisfy the following additional conditions related to axioms of BGI_n:

TA6	if $\mathrm{BEL}(i,\varphi) \in L(w)$, then either $\varphi \in L(w)$ or there exists $v \in W$ such that $(w,v) \in B_i$,
TA45	if $(w,v) \in B_i$ then $\mathrm{BEL}(i,\varphi) \in L(w)$ iff $\mathrm{BEL}(i,\varphi) \in L(v)$,
TA78$_{GB}$	if $(w,v) \in B_i$ then $\mathrm{GOAL}(i,\varphi) \in L(w)$ iff $\mathrm{GOAL}(i,\varphi) \in L(v)$,
TA6$_I$	if $\mathrm{INT}(i,\varphi) \in L(w)$, then either $\varphi \in L(w)$ or there exists $v \in W$ such that $(w,v) \in I_i$,
TA78$_{IB}$	if $(w,v) \in B_i$ then $\mathrm{INT}(i,\varphi) \in L(w)$ iff $\mathrm{INT}(i,\varphi) \in L(v)$,
TA9$_{IG}$	if $(w,v) \in G_i$ and $\mathrm{INT}(i,\varphi) \in L(w)$ then $\varphi \in L(v)$,

Condition **TA6** corresponds to belief consistency[2], **TA45** to positive and negative introspection of beliefs[3], **TA78**$_{GB}$ to positive and negative introspection of goals, **TA78**$_{IB}$ to positive and negative introspection of intentions, and **TA9**$_{IG}$ to goal–intention compatibility.

Given a formula φ we say that $\mathcal{T} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, L)$ is a BGI_n tableau for φ if $\mathcal{T}$ is a a BGI_n tableau and there is a state $w \in W$ such that $\varphi \in L(w)$.

The following analogue to the proposition shown in [9] for $S5_n$ can be shown, giving a basis for an algorithm checking BGI_n satisfiability:

2 This is a condition that occurs in [9], corresponding to the consistency axiom.

3 We give this condition instead of two other conditions given in [9] as correspondents to positive and negative introspection axioms in a $KD45_n$ tableau. The given condition is exactly the condition the authors of [9] give, together with a condition corresponding to the truth axiom for $S5_n$.

Proposition 1 *A formula φ is BGI_n satisfiable iff there is a BGI_n tableau for φ.*

Proof The proof is very similar to the proof given in [9] for $S5_n$ tableaus. Although we have to deal with new conditions here, this is not a problem due to the similarity of conditions **TA78**$_{GB}$, **TA78**$_{IB}$ to condition **TA45**. The direction from left to right is a straightforward adaptation of the proof in [9], and we leave it to the reader.

When constructing a model for φ out of a tableau for φ in the right to left part we have to construct a “serial closure” of some relations. This is done by making isolated states accessible from themselves. For example accessibility relations I'_i for intentions would be defined on the basis of relations I_i in a tableau as follows: $I'_i = I''_i \cup \{(w,w) : \forall v \in W.\ (w,v) \notin I''_i\}$, where I''_i is the smallest set containing I_i and satisfying properties corresponding to axioms **A7**$_{IB}$ and **A8**$_{IB}$. ■

3.1. The algorithm for satisfiability of BGI_n

The algorithm presented below tries to construct, for a given formula φ, a *pre-tableau* – a tree-like structure that forms the basis for a BGI_n tableau for φ. Nodes of this tableau are labeled with subsets of $\neg\mathrm{Sub}(\varphi)$. We give a brief description of the algorithm (see [9] for details) presenting modifications corresponding to new properties in more details:

Input: A formula φ

Step 1 Construct a tree consisting of single node w, with $L(w) = \{\varphi\}$.

Step 2 Repeat until none of the steps **2.1 – 2.2** applies:

Step 2.1 Expand leaves of the tree until they form a propositional tableau or are blatantly inconsistent (nodes that are fully expanded propositional tableaus are called *states* and all other nodes are called *internal nodes*).

Step 2.2 Create successors of all states that are not blatantly inconsistent according to the following rules: (s *denotes a considered state*)

- **bel1** If $\mathrm{BEL}(i,\psi) \in L(s)$ and there are no formulas of the form $\neg\mathrm{BEL}(i,\chi) \in L(s)$, then let $L^{\mathrm{BEL}_i}(s) = \{\chi : \mathrm{BEL}(i,\chi) \in L(s)\} \cup \{\mathrm{OP}(i,\chi) : \mathrm{OP}(i,\chi) \in L(s)\}$, where $\mathrm{OP} \in \{\mathrm{BEL}, \neg\mathrm{BEL}, \mathrm{GOAL}, \neg\mathrm{GOAL}, \mathrm{INT}, \neg\mathrm{INT}\}$. If there is no $\mathrm{b_i}$-ancestor t of s (including $t = s$), such that $L(t) = L^{\mathrm{BEL}_i}(s)$, then create a successor u of s (called $\mathrm{b_i}$–successor) with $L(u) = L^{\mathrm{BEL}_i}(s)$.
- **bel2** If $\neg\mathrm{BEL}(i,\psi) \in L(s)$ then let $L^{\neg\mathrm{BEL}_i}(s) = \{\neg\psi\} \cup L^{\mathrm{BEL}_i}(s)$. If there is no $\mathrm{b_i}$-ancestor t of s (including $t = s$), such that $L(t) = L^{\neg\mathrm{BEL}_i}(s)$, then create a successor u of s (called $\mathrm{b_i}$–successor) with $L(u) = L^{\neg\mathrm{BEL}_i}(s)$.

int1 If $\mathrm{INT}(i,\psi) \in L(s)$ and there are no formulas of the form $\neg\mathrm{INT}(i,\chi) \in L(s)$, then create a successor u of s (called $\mathrm{i_i}$–successor) with $L(u) = L^{\mathrm{INT}_i}(s) = \{\chi : \mathrm{INT}(i,\chi) \in L(s)\}$.

int2 If $\neg\mathrm{INT}(i,\psi) \in L(s)$ then create a successor u of s (called $\mathrm{i_i}$–successor) with $L(u) = L^{\neg\mathrm{INT}_i}(s) = \{\neg\psi\} \cup L^{\mathrm{INT}_i}(s)$.

goal If $\neg\mathrm{GOAL}(i,\psi) \in L(s)$ then create a successor u of s (called $\mathrm{g_i}$–successor) with $L(u) = L^{\neg\mathrm{GOAL}_i}(s) = \{\neg\psi\} \cup \{\chi : \mathrm{GOAL}(i,\chi) \in L(s)\} \cup L^{\mathrm{INT}_i}(s)$.

Step 2.3 Mark a hitherto unmarked node 'satisfiable' if either it is a not blatantly inconsistent state and step **2.2** can not be applied to it and all its successors are marked 'satisfiable', or it is an internal node having at least one descendant marked 'satisfiable'.

Step 3 If the root is marked 'satisfiable' return 'satisfiable', otherwise return 'unsatisfiable'.

Lemma 1 *For any formula φ the algorithm terminates.*

Proof By similar arguments to that in [9] it can be shown that the depth of the pre-tableau constructed by the algorithm is at most $|\varphi|^4$, and so the algorithm must terminate. Note that the depth of the tree is, as in the case of a $S5_n$ tableau, affected mainly by formulas of the form $\mathrm{BEL}(i,\psi)$ and $\neg\mathrm{BEL}(i,\psi)$ in $\neg\mathrm{Sub}(\varphi)$. ■

Lemma 2 *A formula φ is satisfiable iff the algorithm returns 'satisfiable' on input φ.*

Proof For the right to left direction, a tableau $\mathcal{T} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, L)$ based on the pre-tableau constructed by the algorithm is constructed. W is the set of states of the pre-tableau. For $\{w,v\} \subseteq W$, let $(w,v) \in B_i^*$ if v is the closest descendant state of w and the first successor of w on the path between w and v is a $\mathrm{b_i}$–successor of w. Then B_i is determined as the transitive euclidean closure of the above relation B_i^*.

Relations G_i and I_i are defined analogically. Labels of states in W are the same as in the pre-tableau. Checking that $\mathcal{T}$ is a BGI_n tableau is very much like in the case of $S5_n$ tableau, with the new conditions **TA6**, **TA45**, **TA78**$_{GB}$, **TA6**$_I$, **TA78**$_{IB}$ being the most difficult cases.

For **TA6** note that if $v \in W$ has no successor states and $\mathrm{BEL}(i,\psi) \in L(v)$ then v can not be a root (otherwise there is no ancestor of v such that its label is $L^{\mathrm{BEL}_i}(v)$, so step **2.2.bel1** of the algorithm applies to v and it can not be a leaf). If so, there is $w \in W$, such that $(w,v) \in B_i$. Since $\mathrm{BEL}(i,\psi)$ is a subformula of φ, then either $\neg\mathrm{BEL}(i,\psi) \in L(w)$ or $\mathrm{BEL}(i,\psi) \in L(w)$. Because the first possibility leads to contradiction with $\mathrm{BEL}(i,\psi) \in L(v)$, then it must be the second, and this implies $\psi \in L(v)$.

Condition **TA**$_I$ can be shown similarly.

Condition **TA45** is also based on the fact that labels of states are fully expanded propositional tableaus, and can be shown similarly to **TA6** (see [9]). Since **TA78**$_{GB}$ and **TA78**$_{IB}$ are very similar to **TA45**, a $\mathrm{b_i}$–successor inherits all formulas of the form $\mathrm{GOAL}(i,\psi)$, $\neg\mathrm{GOAL}(i,\psi)$, $\mathrm{INT}(i,\psi)$ and $\neg\mathrm{INT}(i,\psi)$, then they can be shown analogically to **TA45**. **Proposition 1** gives the final result.

For the left to right direction we show, for any node w in the pre-tableau, the claim that if w is not marked 'satisfiable' then $L(w)$ is inconsistent. From this it follows that if the root is not marked 'satisfiable' then $\neg\varphi$ is provable and thus φ is unsatisfiable.

The claim is shown by induction on the length of the longest path from a node w to a leaf of the pre-tableau. Most cases are easy and can be shown similarly to the case of $S5_n$ presented in [9]. We show only the most difficult case connected with new axioms of BGI_n, namely the one, in which w is not a leaf and has a $\mathrm{b_i}$–successor v generated by a formula of the form $\mathrm{BEL}(i,\psi) \in L(w)$ (other cases are either similar or easier). Since by induction hypothesis $L(v)$ is inconsistent, we can show using **A2**, **R1** and **R2** that the set $X = \{\mathrm{BEL}(i,\psi) : \mathrm{BEL}(i,\psi) \in L(w)\}) \cup \{\mathrm{BEL}(i,\psi) : \psi \in L(w) \text{ and is of the form } \mathrm{OP}(i,\chi)\}$ proves $\mathrm{BEL}(i,\bot)$, so by **A6** X is also inconsistent. Now assume that $L(w)$ is consistent, then the set $Y = L(w) \cup \{\mathrm{BEL}(i,\psi) : \psi \in L(w) \text{ and is of the form } \mathrm{OP}(i,\chi)\} \cup \{\neg\mathrm{BEL}(i,\bot)\}$ is also consistent (by axioms **A4-6**, **A7-8**$_G B$ and **A7-8**$_I B$). This leads to contradiction, since $X \subseteq Y$, and thus $L(w)$ must be inconsistent. ■

Since the depth of the pre-tableau constructed by the algorithm for a given φ is at most $|\varphi|^4$, and the algorithm is deterministic, it can be run on a deterministic Turing machine by depth-first search using polynomial space. Thus BGI_n is in PSPACE. On the other hand the problem of KD_n satisfiability, known to be PSPACE-hard, can be reduced to BGI_n satisfiability, so BGI_n is PSPACE–complete.

4. Complexity of the system $BGI_n^{C,M}$

We will show that the satisfiability problem for a $BGI_n^{C,M}$ system is EXPTIME–complete. First we prove that $BGI_n^{C,M}$ has a small model property. To show this a filtration technique is used (see [2]). Let $G \subseteq \{1,\ldots,n\}$.

A set of formulas Σ closed for subformulas is *closed* if it satisfies the following:

Cl1 if $\text{C-BEL}_G(\varphi) \in \Sigma$, then $\text{E-BEL}_G(\varphi \wedge \text{C-BEL}_G(\varphi)) \in \Sigma$,

Cl2 if $\text{E-BEL}_G(\varphi) \in \Sigma$, then $\{\mathrm{BEL}(j,\varphi) : j \in G\} \subseteq \Sigma$,

Cl3 if $\text{M-INT}_G(\varphi) \in \Sigma$, then $\text{E-INT}_G(\varphi \wedge \text{M-INT}_G(\varphi)) \in \Sigma$,

Cl4 if $\text{E-INT}_G(\varphi) \in \Sigma$, then $\{\mathrm{INT}(j,\varphi) : j \in G\} \subseteq \Sigma$.

Let $\mathcal{M} = (W, \{B_i : i \in \mathcal{A}\}, \{G_i : i \in \mathcal{A}\}, \{I_i : i \in \mathcal{A}\}, Val)$ be a BGI_n model, Σ a closed set, and let $\equiv_f^\Sigma \subseteq \Sigma \times \Sigma$ be an equivalence relation such that, for $\{w,v\} \subseteq W$, $w \equiv_f^\Sigma v$ iff for any $\varphi \in \Sigma$, $\mathcal{M}, w \models \varphi \Leftrightarrow \mathcal{M}, w \models \varphi$.

Let $\mathcal{M}^f_\Sigma = (W^f, \{B^f_i : i \in \mathcal{A}\}, \{G^f_i : i \in \mathcal{A}\}, \{I^f_i : i \in \mathcal{A}\}, Val^f)$ be defined as follows:

F0 $W^f = W/\equiv^\Sigma_f, Val^f(p,[w]) = Val(p,w)$,

F1 $B^f_i = \{([w],[v]) :$ for any $\mathrm{BEL}(i,\varphi) \in \Sigma,\ \mathcal{M}, w \models \mathrm{BEL}(i,\varphi) \Rightarrow \mathcal{M}, v \models \varphi$ and for any $\mathrm{OP}(i,\varphi) \in \Sigma,\ \mathcal{M}, w \models \mathrm{OP}(i,\varphi) \Leftrightarrow \mathcal{M}, v \models \mathrm{OP}(i,\varphi)\}$, where $\mathrm{OP} \in \{\mathrm{BEL}, \mathrm{GOAL}, \mathrm{INT}\}$.

F2 $G^f_i = \{([w],[v]) :$ for any $\mathrm{GOAL}(i,\varphi) \in \Sigma,\ \mathcal{M}, w \models \mathrm{GOAL}(i,\varphi) \Rightarrow \mathcal{M}, v \models \varphi$ and for any $\mathrm{INT}(i,\varphi) \in \Sigma,\ \mathcal{M}, w \models \mathrm{INT}(i,\varphi) \Rightarrow \mathcal{M}, v \models \varphi\}$.

F3 $I^f_i = \{([w],[v]) :$ for any $\mathrm{INT}(i,\varphi) \in \Sigma,\ \mathcal{M}, w \models \mathrm{INT}(i,\varphi) \Rightarrow \mathcal{M}, v \models \varphi\}$.

It is easy to check that if $\mathcal{M}$ is a BGI_n model, then so is $\mathcal{M}^f_\Sigma$ and, moreover, that if Σ is a closed set, then $\mathcal{M}^f_\Sigma$ is a *filtration* of $\mathcal{M}$ through Σ. This leads to the following standard lemma (thus left without a proof):

Lemma 3 *If $\mathcal{M}$ is a BGI_n model and Σ is a closed set of formulas then for all $\varphi \in \Sigma$ and all $w \in W$, $\mathcal{M}, w \models \varphi$ iff $\mathcal{M}^f_\Sigma, [w] \models \varphi$.*

From **lemma 3** it follows that $BGI^{C,M}_n$ has the finite model property and that its satisfiability problem is decidable. Let $\mathrm{Cl}(\varphi)$ denote the smallest closed set containing $\mathrm{Sub}(\varphi)$, and let $\neg\mathrm{Cl}(\varphi)$ consist of all formulas in $\mathrm{Cl}(\varphi)$ and their negations. If a formula φ is satisfiable then it is satisfiable in a filtration through $\mathrm{Cl}(\varphi)$, and any such filtration has at most $|P(\mathrm{Cl}(\varphi))| = \mathcal{O}(2^{|\varphi|})$ states.

Now we present an exponential time algorithm for checking $BGI^{C,M}_n$ satisfiability of a formula φ. The algorithm and the proof of its validity are modified versions of the algorithm for checking satisfiability for PDL and its validity proof presented in [10][4]. The algorithm attempts to construct a model $\mathcal{M} = \mathcal{N}^f_{\mathrm{Cl}(\varphi)}$, where $\mathcal{N}$ is a *canonical model* for $BGI^{C,M}_n$. This is done by constructing a sequence of models $\mathcal{M}^k$, being subsequent approximations of $\mathcal{M}$ as follows:

Input: A formula φ

Step 1 Construct a model $\mathcal{M}^0 = (W^0, \{B^0_i : i \in \mathcal{A}\}, \{G^0_i : i \in \mathcal{A}\}, \{I^0_i : i \in \mathcal{A}\}, Val^0)$, where W^0 is the set of all maximal subsets of $\neg\mathrm{Cl}(\varphi)$ (that is sets that for every $\psi \in \mathrm{Cl}(\varphi)$ contain either ψ or $\neg\psi$, $Val^0(p,w) = 1$ iff $p \in w$, and accessibility relations are defined analogically as in $\mathcal{M}^f_{\mathrm{Cl}(\varphi)}$. We present the definition of B^0_i, which makes definitions **G1**, **I1** of G^0_i and I^0_i obvious:

B1 $B^0_i = \{(w,v) :$ for any $\mathrm{BEL}(i,\varphi) \in \mathrm{Cl}(\varphi)$, $\mathrm{BEL}(i,\varphi) \in w \Rightarrow \varphi \in v$ and for any $\mathrm{OP}(i,\varphi) \in \mathrm{Cl}(\varphi),\ \mathrm{OP}(i,\varphi) \in w \Leftrightarrow \mathrm{OP}(i,\varphi) \in v\}$, where $\mathrm{OP} \in \{\mathrm{BEL}, \mathrm{GOAL}, \mathrm{INT}\}$.

4 Note that one can see the $BGI^{C,M}_n$ logic as a modified and restricted version of PDL, where BEL, GOAL and INT operators for each agent are seen as atomic programs satisfying some additional axioms, while group operators can be defined as complex programs using the $\cup$ and *.

Step 2 Construct a model $\mathcal{M}^1$ by removing from W^0 states that are not closed propositional tableaus.

Step 3 Repeat the following (starting with $k = 0$ until no state can be removed:

Step 3.1 Find a formula $\psi \in \neg\mathrm{Cl}(\varphi)$ and state $w \in W^k$ such that $\psi \in w$ and one of the conditions[5]. below is not satisfied. It the state was found, remove it from W^k to obtain W^{k+1}.

AB1 if $\psi = \neg\mathrm{BEL}(i,\chi)$, then there exists $v \in B^k_i$ such that $\neg\chi \in v$ (analogical conditions **AG1** and **AI1** for GOAL and INT,

AB2 if $\psi = \mathrm{BEL}(i,\chi)$, then there exists $v \in B^k_i$ such that $\chi \in v$ (analogical condition **AI2** for INT,

AEB1 if $\psi = \neg\text{E-BEL}_G(i,\chi)$, then there exists $v \in B^k_G$ (where $B^k_G = \bigcup_{j \in G} B^k_j$ such that $\neg\chi \in v$ (analogical condition **AEI1** for E-INT_G,

ACB1 if $\psi = \neg\text{C-BEL}_G(i,\chi)$, then there exists $v \in (B^k_G)^*$ such that $\neg\chi \in v$ (analogical condition **AMI1** for M-INT_G,

Step 4 If there is a state in the model $\mathcal{M}^l$ obtained after step **3** containing φ, then return 'satisfiable', otherwise return 'unsatisfiable'.

It is obvious that the algorithm terminates. Moreover, since each step can be done in polynomial time, the algorithm terminates after $\mathcal{O}(2^{|\varphi|})$ steps. To prove the validity of the algorithm, we have to prove an analogue to a lemma in [10]: ($\mathrm{OP}_G \in \{\text{E-BEL}_G, \text{E-INT}_G\}$, $\mathrm{OP}^*_G \in \{\text{C-BEL}_G, \text{M-INT}_G\}$ *and R denotes the relation corresponding to operator* OP *used in the particular context*)

Lemma 4 *Let $k \geq 1$ and assume that $\mathcal{M} \subseteq \mathcal{M}^k$. Let $\chi \in \mathrm{Cl}(\varphi)$ be such that every formula from $\mathrm{Cl}(\chi)$ of the form $\mathrm{OP}(i,\psi)$, $\mathrm{OP}_G(\psi)$ or $\mathrm{OP}^*_G(\psi)$ and $w \in W^k$ satisfies the conditions of step **3** of the algorithm. Then:*

1 *for all $\xi \in \mathrm{Cl}(\chi)$ and $v \in W^k$, $\xi \in v$ iff $\mathcal{M}, v \models \xi$;*

2.1 *for any $\mathrm{OP}(i,\xi) \in \mathrm{Cl}(\chi)$ and $\{w,v\} \subseteq W^k$:*

2.1.a *if $(w,v) \in R_i$ then $(w,v) \in R^k_i$;*

2.1.b *if $(w,v) \in R^k_i$ and $\mathrm{OP}(i,\xi) \in v$ then $\xi \in v$;*

2.2 *for any $\mathrm{OP}_G(\xi) \in \mathrm{Cl}(\chi)$ and $\{w,v\} \subseteq W^k$:*

2.2.a *if $(w,v) \in R_G$ then $(w,v) \in R^k_G$;*

2.2.b *if $(w,v) \in R^k_G$ and $\mathrm{OP}_G(\xi) \in v$ then $\xi \in v$;*

2.3 *for any $\mathrm{OP}^*_G(\xi) \in \mathrm{Cl}(\chi)$ and $\{w,v\} \subseteq W^k$:*

2.3.a *if $(w,v) \in (R_G)^*$ then $(w,v) \in (R^k_G)^*$;*

2.3.b *if $(w,v) \in (R^k_G)^*$ and $\mathrm{OP}^*_G(\xi) \in v$ then $\xi \in v$.*

Proof The proof is analogical to the one of the lemma for PDL and the additional properties of $BGI^{C,M}_n$ do not affect the argumentation. The proof of points **2.1-3** is essentially based on the fact that $\mathcal{M}$ is a filtration and similar techniques are used here to those from the proof of the filtration lemma. The proof of point **1** is by induction on the structure of ξ, similarly to its analogue for the lemma for PDL. ■

5 The conditions are analogical to conditions for PDL. The only difference are conditions **AB2** and **AI2** that correspond to axioms **A6** and **A6**$_I$

Since every state $w \in W$ is a maximal subset of $\neg\mathrm{Cl}(\varphi)$, we have $W \subseteq W^0$. Moreover, since every state $w \in W$ is a propositional tableau satisfying conditions from the step **2** of the algorithm, thus $W \subseteq W^1$. Conditions in step **2** guarantee also that no state $w \in W$ can be deleted in step **3**. This shows that $W \subseteq W^k$, for all W^k constructed throughout an execution of the algorithm. It follows that if φ is satisfiable then the algorithm will return 'satisfiable'.

If model $\mathcal{M}^l$ obtained after step **3** of the algorithm is not empty, then it can be easily checked that it is a BGI_n model. This is because every model $\mathcal{M}^k$ constructed throughout an execution of the algorithm preserves conditions **B1, G1, I1**. Moreover, conditions **AB2**, **AI2** guarantee that relations B_i^l and I_i^l are serial.

Now, if there is a $w \in W^l$ such that $\varphi \in w$, then (by **1** of **lemma 4**) $\mathcal{M}^l, w \models \varphi$, and since M^l is a BGI_n model, then φ is $BGI_n^{C,M}$ satisfiable. So the algorithm is valid.

Remark We kept the algorithm above and **lemma 4** similar to the ones presented in [10], so one can combine them to obtain a deterministic exponential time algorithm for a combination of $BGI_n^{C,M}$ and PDL.

5. Discussion and conclusions

This paper deals with the complexity of two important components of our theory of teamwork. The first covers agents' individual attitudes, including their interdependencies, which makes the decision procedure more complicated. The second one deals with the team attitude par excellence: collective intention. Importantly, however, our results have a more general impact. The tableau methods used in this paper can be adapted to the non-temporal parts of other multi-modal logics which are similar in spirit to ours, such as the KARO framework [1].

The presented system defining collective intentions is decidable: in this paper, we have proved to be EXPTIME-complete, so in general it is feasible to give automated proofs of desired properties only for formulas of limited length. As with other modal logics, the better option would be to develop a variety of different algorithms and heuristics, each performing well on a limited class of inputs. For example, it is known that restricting the number of propositional atoms to be used and/or the depth of modal nesting may reduce the complexity (cf. [8, 11]). Also, when considering specific applications it is possible to reduce some of the infinitary character of collective beliefs and intentions to more manageable proportions (cf. [7, Ch. 11]). It is particularly interesting to restrict to formulas in Horn form, which seems natural not only in MAS but also more generally in AI applications [13]. We plan to investigate these and other reductions in a forthcoming paper. Such reductions in complexity are essential when the strongest motivational attitude, collective commitment, is considered [6].

References

[1] H. Aldewereld, W. van der Hoek, and J.-J. Meyer. Rational teams: Logical aspects of multi-agent systems. *Fundamenta Informaticae*, 63 (2,3):159–183, 2004.

[2] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2002.

[3] P. Blackburn and E. Spaan. A modal perspective on the computational complexity of attribute value grammar. *Journal of Logic, Language and Information*, 2:129–169, 1993.

[4] B. Dunin-Kęplicz and R. Verbrugge. Collective intentions. *Fundamenta Informaticae*, 51(3):271–295, 2002.

[5] B. Dunin-Kęplicz and R. Verbrugge. Evolution of collective commitments during teamwork. *Fundamenta Informaticae*, 56:329–371, 2003.

[6] B. Dunin-Kęplicz and R. Verbrugge. A tuning machine for cooperative problem solving. *Fundamenta Informaticae*, 63:283–307, 2004.

[7] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, MA, 1995.

[8] J. Halpern. The effect of bounding the number of primitive propositions and the depth of nesting on the complexity of modal logic. *Artificial Intelligence*, 75:361–372, 1995.

[9] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54(3):319–379, 1992.

[10] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic. Foundations of Computing*. MIT Press, Cambridge, 2000.

[11] U. Hustadt and R. Schmidt. On evaluating decision procedures for modal logics. In M. Pollack, editor, *Proceedings IJCAI'97*, Los Angeles (CA), 1997. Morgan Kauffman.

[12] M. Kacprzak, A. Lomuscio, and W. Penczek. From bounded to unbounded model checking for interpreted systems. *Fundamenta Informaticae*, 63 (2,3):107–308, 2004.

[13] L. A. Nguyen. On the complexity of fragments of modal logics. In R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing, editors, *Preproceedings of AIML 2004*, pages 318–330, Manchester, 2004. University of Manchester.

[14] J. van Benthem. Correspondence theory. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 167–247. Reidel, Dordrecht, 1984.