Combinations of Normal and Non-normal Modal Logics for Modeling Collective Trust in Normative MAS

Clara Smith[1,2], Agustín Ambrossio[1], Leandro Mendoza[2], and Antonino Rotolo[3]

[1] FACET, UCALP, Argentina
[2] Facultad de Informática, UNLP, Argentina
[3] CIRSFID, University of Bologna, Italy

Abstract. We provide technical details for combining normal and a non-normal logics for the notion of collective trust. Such combinations lead to different levels of expressiveness of the system. We give a possible structure for a combined model checker for one of the logic resulting from such combinations.

1 Motivation and Aims

Trust protection plays an important role in the law [15,21,18]. Such a protection is often and typically related to the problem of providing tools to support legally valid interactions between any kind of agents and/or to legally ground contractual transactions [15,21]. Indeed, trust protection is strongly implemented especially when agents' beliefs seem reasonable or when trustees' behaviour induces trusters' reliance. However, in multi-lateral agreement it is often the case that such reliance is mutual and this fact is relevant for trust protection. In particular, if agent x breaks group trust with regard to A, trust deception must be checked against the fact that x was supposed by the others to intend A, and x believed so.

Any computer application providing tools for detecting collective trust deception in the legal domain should require to develop

– a sound and rigorous formal analysis of the notion of collective trust,
– reasoning methods for computing when collective trust emerges and occurs in an arbitrarily large group of agents.

This general aim of this paper is to contribute to the above two research issues by significantly extending [19]'s results. In [19] Smith and Rotolo adopted [8]'s cognitive model of *individual* trust in terms of necessary mental ingredients which settle under what circumstances an agent x trusts another agent y with regard to an action or state-of-affairs, i.e. under which beliefs and goals an agent delegates a task to another agent. Using this characterization of individual trust, these authors provided a logical reconstruction of different types of *collective* trust, which for example emerge in groups with multilateral agreement, or which are the glue for grounding *in solidum* obligations raising from a "common front" of agents (for example, each member of the front can behave, in principle, as creditor or debtor of the whole). These collective cognitive states were characterized in [19] within a multi-modal logic based on [3]'s axiomatization for collective beliefs and intentions combined with a non-normal modal logic for the operator Does for agency. Such a combination was based on the following assumptions:

M. Palmirani et al. (Eds.): AICOL Workshops 2011, LNAI 7639, pp. 189–203, 2012.

Observation 1 (Expressiveness of the system). *A formula like* $\mathrm{Does}_i\mathscr{A}$ *means that "agent i brings it about that* $\mathscr{A}$*". In this setting, the* Does *modalities are always applied to atomic propositional constants representing single behavioural actions, as in e.g.* Does_x *PayBill (which is meant to stand for "agent x pays the bill"). In the theory under study, normal operators interact with the* Does *modality in a restricted one-way manner: agents' actions always appear as innermost operators within well-formed formulas, as in e.g.* $\mathrm{Bel}_y(\mathrm{Does}_x\ \mathit{PayBill})$ *(which is meant to stand for "the agent y believes that agent x pays the bill"). This means that no modality can occur in the scope of* Does.

Observation 2 (Semantics). *[19, Definition 2] proposed for the above mentioned system a semantics embedding standard multi-modal Kripke semantics for mental states into a Scott-Montague (multi-relational) semantics for* Does *[11].*

These combination and semantic embedding were assumed correct because Kripke semantics can be seen as a special case of multi-relational semantics. Although the referred concrete embedding appears to be straightforward in [19], it is worth pointing out that some basic results such as completeness and decidability are not immediately obvious and require some detailed technical machinery. This paper fills this gap by showing a simple way to prove those results, and also describes a model checking algorithm for that logic: the possibility of designing a model checker indicates that such logic can provide a feasible interpretation for norm-governed multi-agent systems and a method for computing collective trust.

The rest of the paper is organized as follows. Section 2 presents the main concepts of [19] and the proposed logical system. Section 3 reorganizes the multi-relational model in [19] as a particular combination of modal logics, which amounts to place the normal logics on top of the non-normal logics. For doing this, we first obtain two restrictions of the original logics. By exploiting results in regard to some techniques for combining logics, we prove that [19]'s system is complete and decidable. Hence, the sketch for an appropriate model checker is also outlined. Section 5 presents an independent combination of the normal and the non-normal counterparts of the base logics. This combination leads to an ontology of pairs of state-of-affairs which allows a structural basis for more expressiveness. For example, it is possible to write and test in the new ontology formulas such as $\mathrm{Does}_i\ (\mathrm{Does}_j\ (\mathrm{Goal}\ \mathscr{A}))$. Some brief conclusions end the paper.

2 Background

There are situations where complex collective patterns are involved in social and legal interaction. Suppose that three agents x, y and z agree that some goal A should be jointly achieved. Some kind of coordination among them is of course required, but, minimally, such a multi-lateral agreement at least implies that each agent trusts that the others jointly intend to achieve A, and also believe in that. This simple agreement thus presumes a relatively elaborated collective trust background.

Collective trust and the corresponding delegation of tasks can be weak or strong [8]: weak delegation means that there are delegation situations which do not suppose any agreement, deal or promise at all, nor which yield to rights; strong delegation are the basis for promises, commitments and conventions. Since these forms of delegation support different degrees of trust intensity, different corresponding types of collective trust—joint trust, reliance, and collective trust—were introduced in [19] and can be illustrated as follows:

Example 1 (Joint trust). Suppose that agent y is at the bus stop, and there is a group G of people standing not at the bus stop but close to y, *expecting* that y will raise her hand and stop the bus.

Example 2 (Reliance). It is Mary's birthday. Her co-workers give some money to y, another co-worker who is going downtown, *relying* on y for the search and purchase of a gift. Everyone *trusts* that y will do so.

Example 3 (Collective trust). Student bands build up street-puppets filled with fireworks, which are to be burned on New Year's day. Each band builds its chosen puppet-of-the-year from scratch. The town administration institutionalized a competition and settled an award for the best figure. Bands' custom establishes that figures ought to be watched and protected day and night, this because a very common practice is to burn other bands' figures before the New Year's day by sending one band member (a saboteur). The consequence of successful sabotages is the exclusion of opponents from the competition. Assume the student band G *entrusts* its member s to burn H's puppet.

Accordingly, joint trust simply consists in the fact that all individuals in a group trust another agent for achieving their goal, reliance requires some mutual intentional coordination within the group, and collective trust assumes that the group is aware of such a coordinating effort to achieve a goal.

The following subsection outlines [19]'s logical framework, which will be our starting point for the subsequent sections.

2.1 The Logical Framework

The multi-modal language of [19] works with a finite set of agents $A = \{x, y, z, ...\}$ and a countable set of atomic propositional sentences usually denoted by $P = \{p, q, r, ...\}$. Complex expressions are formed syntactically from these in the usual inductive way using $\perp$ (*false*) and $\top$ (*true*), standard Boolean connectives, and the unary modalities we describe next.

The operator $\mathrm{Goal}_x\ \mathscr{A}$ is used to mean that "agent x has goal $\mathscr{A}$", where $\mathscr{A}$ is a proposition. Propositions reflect particular state-of-affairs, as in [3]. $\mathrm{Int}_x\ \mathscr{A}$ is meant to stand for "agent x has the intention to make $\mathscr{A}$ true". Intentions within the area of Co-operative Problem Solving (*CPS*) are viewed as inspiration for goal-directed activities. The doxastic (or epistemic) modality $\mathrm{Bel}_x\ \mathscr{A}$ represents that "agent x has the belief that $\mathscr{A}$". The $\mathrm{Does}_x\ \mathscr{A}$ operator is to be understood in the same sense given in Elgesem's account to represent successful agency, i.e. "x brings it about that $\mathscr{A}$" [5]. To simplify technicalities, the logic in [19] assumes that in expressions like $\mathrm{Does}_x\ \mathscr{A}$ no modal operators occur in the scope of Does; therefore $\mathscr{A}$ denotes any behavioral action concerning a conduct, such as withdrawal, inform, purchase, payment, etc. We will assume the same restriction for Section 3, and we will eliminate it in Section 5 for regaining expressiveness.

As classically established [3], Goal is a K_n operator, while Int and Bel are, respectively, KD_n and $KD45_n$. The logic of Does, instead, is non-normal, it is closed under logical equivalence and amounts to the following schemata [5,11]: $\mathrm{Does}_x\ \mathscr{A} \rightarrow \mathscr{A}$, $(\mathrm{Does}_x\ \mathscr{A} \wedge\ \mathrm{Does}_x\ \mathscr{B}) \rightarrow \mathrm{Does}_x(\mathscr{A} \wedge \mathscr{B})$, $\neg\mathrm{Does}_x\top$, and $\neg\mathrm{Does}_x\bot$.

Remark 1. The main difference between [19]'s logic (let us call it $\mathfrak{F}$) and [3]'s system is that $\mathfrak{F}$ embeds Does and introduces new (non-primitive) operators defined on the basis of the [3]'s ones. First, $\mathfrak{F}$ defines the single-agent trust operator Trust (an agent x trusts another agent y with respect to a state of affairs ϕ) as follows:

$$\mathrm{Trust}_x^y\phi \equiv \mathrm{Goal}_x\phi \wedge \mathrm{Bel}_x\mathrm{Does}_y\phi \wedge \mathrm{Int}_x(\mathrm{Does}_y\phi \wedge \neg\mathrm{Does}_x\phi) \wedge \mathrm{Goal}_x\mathrm{Int}_y\phi \wedge \mathrm{Bel}_x\mathrm{Int}_y\phi \quad (1)$$

If G is a group of agents, the other derived operators of [19] are introduced to capture joint trust, reliance, and collective trust, respectively (see Examples 1, 2, and 3):

$$\mathrm{JTrust}_y^G A \equiv (\bigwedge_{i\in G} \mathrm{Trust}_y^i A) \quad (2)$$

$$\mathrm{Rel}_y^G A \equiv \mathrm{JTrust}_y^G A \wedge \mathrm{MInt}_G(\mathrm{JTrust}_y^G A) \quad (3)$$

$$\mathrm{CTrust}_s^G A \equiv \mathrm{Rel}_s^G A \wedge \mathrm{CBel}_G(\mathrm{Rel}_s^G A) \quad (4)$$

where the axiomatizations for MInt (mutual intention) and CBel (common belief) are those proposed in [3]:

$$\mathrm{MInt}^G A \equiv (\textstyle\bigwedge_{i\in G} \mathrm{Int}^i(A \wedge \mathrm{MInt}^G A)) \qquad \mathrm{CBel}^G A \equiv (\textstyle\bigwedge_{i\in G} \mathrm{Bel}^i(A \wedge \mathrm{CBel}^G A))$$

3 Combining the Logics by Modalization/Temporalization

In this section we show how to characterize the logic of [19] as the combination of the component logics (the logic of Does and the normal component of [19]'s system) using the so-called temporalization/modalization techniques.

Before reorganizing the logic of [19] in this way, we recall some background knowledge. As is well-known, Scott-Montague semantics is a generalization of the traditional Kripke semantics [12]. Instead of a collection of worlds connected to a given world w through a relation R, consider a set of collections of worlds connected to w. These collections are the *neighbourhoods* of w. Formally, a Scott-Montague frame is an ordered pair $\langle W, N\rangle$ where W is a set of worlds and N is a function assigning to each w in W a set of subsets of W (the neighbourhoods of w). A Scott-Montague model is a triple $\langle W, N, V\rangle$ where $\langle W, N\rangle$ is a Scott-Montague frame and V is a valuation function defined as for Kripke frames, except for $\Box\mathscr{A}$: it is true at w iff the set of elements of W where $\mathscr{A}$ is true is one of the sets in $N(w)$; i.e., iff it is a neighbourhood of w.

Let us bring in the structure discussed in [19]. It is a multi-relational frame of the form [11]:

$$\mathfrak{F} = \langle A, W, \{B_i\}_{i\in A}, \{G_i\}_{i\in A}, \{I_i\}_{i\in A}, \{D_i\}_{i\in A} \rangle$$

where:

- A is the finite set of agents;
- W is a set of situations, or points, or possible worlds;
- $\{B_i\}_{i\in A}$ is a set of accessibility relations wrt Bel, which are transitive, euclidean and serial;
- $\{G_i\}_{i\in A}$ is a set of accessibility relations wrt Goal, (standard K_n semantics);
- $\{I_i\}_{i\in A}$ is a set of accessibility relations wrt Int, which are serial; and
- $\{D_i\}_{i\in A}$ is a family of sets of accessibility relations D_i wrt Does, which are pointwise closed under intersection, reflexive and serial [11].

A model based on $\mathfrak{F}$ is in its turn of the form $\langle \mathfrak{F}, V \rangle$, where V is the corresponding valuation function ([19, Definition 2]). Notice that [11] proved that Scott-Montague and multi-relational semantics are equivalent for the propositional case, so they can be interchangeably used.

Put this way, it is easy to identify two overlapping “nets" of relations over the same set W. The first net (or multi-graph) corresponds to “wires" for normal operators, the second net corresponds to the accessibility relations for the Does_i modalities [1].

Following, we can assert two facts based on Definition 4.24 and Theorem 4.22 in [2] (which respectively settle how to construct a canonical model for a normal logic, and state that a normal modal logic is strongly complete with respect to its canonical model). First, that the modal similarity type built up from the normal modalities above has a canonical model; second, that this logic is complete w.r.t. its canonical model. Let us call **N** the logic with signature (Bel, Int, Goal) above (the normal modalities); hence **N** is a normal multi-modal multi-agent logic, which is complete (this proof is available in [1], we also sketch it in the Appendix).

Taking into account Observation 1 and what was stated regarding **N**, and according to the definition of temporalization given by Finger and Gabbay [9], (see also [10]) the system in [19] can be seen as a combination of logics where the normal modal machinery is placed on top of the non-normal logic. The non-normal equipment is in its turn multi-modal, as there is one Does_i modality for each agent i. Indeed, [9]'s techniques were originally designed for temporalizing logics and are a special case of the modalization ones [6], which simply use the same intuition with the aim of externally applying *any* (even non-normal) modal logic to *any* generic logic system[2]. The advantage of this approach is that the resulting logic obtained from the combination is complete and decidable if both its components are, too.

Let us develop this insight.

[1] This definition does not include the Obl modality for obligations. Obl was originally incorporated in $\mathfrak{F}$ for dealing with the deontic connotation of an operator of the theory [19, sec. 4]. We will omit it in what follows to keep the set of modalities manageable. We come back to Obl later with the purpose to showing further possibilities for combining logics (Section 6).

[2] It has been very recently proved that modalization/temporalization techniques used in this paper are simple instances of non-iterated asymmetric importing and fibring techniques [17,16].

Consider $\mathfrak{F}$ as a split into an outer normal multi-modal frame, and inner Scott-Montague frames.

Provided this rearrangement, the intuition behind the valuation of formulas within the system is the following. When we evaluate normal operators (e.g. we parse a formula) we navigate through the outer Kripke model. When a Does_i formula at any given point w is to be evaluated, we navigate through a Scott-Montague model.

The following subsection presents the technical aspects of the temporalization/modalization.

3.1 Modalization: Syntax and Semantics

Take the logic in [19]. Call **N** the restriction of $\mathfrak{F}$ to its normal part, and call Does the restriction of $\mathfrak{F}$ to its non-normal part. We can safely assume that Does is a propositional logic [11]. According to the methodology in [10], we partition the set of formulas in Does into two subsets: Boolean formulas, BDoes, and monolithic formulas, MDoes. A formula $\mathscr{A}$ belongs to BDoes if its outermost operator is a Boolean connective (e.g. $\text{Does}_x\, \mathscr{A} \wedge \text{Does}_x\, \mathscr{B}$); otherwise it belongs to MDoes (e.g. $\text{Does}_x\, \mathscr{A}$). It is clear that there is no intersection among the set of modalities of **N** and Does. Call **N**(Does) the modalization of Does by means of **N**.

N(Does): Syntax Let $\mathscr{L}_{\text{Does}}$ denote the language of the logic of agency (with no normal modalities and without their syntax formation rules), and $\mathscr{L}_{\mathbf{N}}$ denote the language of **N** (without the Does modality and its syntax formation rule). The language $\mathscr{L}_{\mathbf{N}(\text{Does})}$ of **N**(Does)—over the set of proposition letters P—is obtained by replacing the formation rule of sentences in $\mathscr{L}_{\mathbf{N}}$ that says "every proposition letter in P is a formula" by the formation rule:

$$every\ monolithic\ formula\ in\ \mathscr{L}_{\text{Does}}\ is\ a\ formula$$

As pointed out in [9], this replacement can be matched with a process called "fuzzling" or layering: formulas in the base system can be substituted for atoms of the top system.

To formally outline the semantics for the modalization, we need a reframing of models based on $\mathfrak{F}$ in terms of the restricted models.

A modalized model for **N**(Does) has the structure:

$$\langle A, W, \{B_i\}_{i\in A}, \{G_i\}_{i\in A}, \{I_i\}_{i\in A}, V', \{d_i\}\rangle$$

where:

- A is a finite set of agents;
- W is a set of points, or possible worlds;
- $\{B_i\}_{i\in A}$ is a set of accessibility relations wrt Bel, which are transitive, euclidean and serial;
- $\{G_i\}_{i\in A}$ is a set of accessibility relations wrt Goal;
- $\{I_i\}_{i\in A}$ is a set of accessibility relations wrt Int, which are serial;
- V' is the valuation function V restricted to the normal operators, defined as follows:

1. standard Boolean conditions;
2. $V'(w, \text{Bel}_i\ \mathscr{A}) = 1$ iff $\forall v \in W(\text{if } wB_iv \text{ then } V'(v,\mathscr{A}) = 1)$;
3. $V'(w, \text{Goal}_i\ \mathscr{A}) = 1$ iff $\forall v \in W(\text{if } wG_iv \text{ then } V'(v,\mathscr{A}) = 1)$;
4. $V'(w, \text{Int}_i\ \mathscr{A}) = 1$ iff $\forall v \in W(\text{if } wI_iv \text{ then } V'(v,\mathscr{A}) = 1)$; and

– each d_i is a total function mapping, for each world w in W, for each agent i, into a multi-relational model of the form:

$$\eta = \langle W, D_i, \mathtt{v} \rangle$$

where:
– W is the (same, original) set of worlds,
– D_i is a family of sets of accessibility relations $\mathtt{D}_i$ wrt agency regarding agent i, which are pointwise closed under intersection, reflexive and serial [11],
– $\mathtt{v}$ is V restricted to the non-normal operators. That is, the valuation function for agency that says that $\text{Does}_i\ \mathscr{A}$ holds in w if and only if the set of worlds where $\mathscr{A}$ is true is one of the neighborhoods of w. Formally:

1. standard Boolean conditions;
2. $\mathtt{v}(w, \text{Does}_i\ \mathscr{A}) = 1$ *iff* $\exists \mathtt{D}_i \in D_i$ *such that* $\forall u(w\mathtt{D}_iu$ *iff* $\mathtt{v}(u,\mathscr{A}) = 1)$.

Let us call $\mathscr{K}\mathscr{L}_{\text{Does}}$ the set of models for $\mathscr{L}_{\text{Does}}$, then d_i: $W \to \mathscr{K}\mathscr{L}_{\text{Does}}$.

The above semantics instantiates the construction criteria of Definition 4.2 in [9] and their generalization in [6], and so corresponds to a case of temporalization/modalization.

N(Does): Semantics Given a model $\mathfrak{M}$, given $w \in W$, given V' valuation function in $\mathfrak{M}$, and given functions d_i, the semantics for **N**(Does) is obtained by replacing the clause for **N** that says

$$\mathfrak{M}, w \models p \textit{ iff } p \in V'(w), \textit{whenever } p \in P$$

with the clause:

$$\mathfrak{M}, w \models \mathscr{A} \textit{ iff } d_i(w) \models \mathscr{A}, \textit{whenever } \mathscr{A} \in \text{MDoes}.$$

Note here that $\mathscr{A}$ has the form $\text{Does}_i\ \mathscr{B}$, as $\mathscr{A}$ is a monolithic formula.

Once a formula has entered the "Does component" it cannot come back to the top level [10]. Accordingly, we cannot test the validity of statements such as $\text{Does}_i(\text{Goal}_j\ \mathscr{A})$ (which can be seen as capturing a form of persuasion: "agent i makes agent j have $\mathscr{A}$ as a goal"). We address a possible solution to this drawback in Section 5.

Notice also that we combine the logics in a rather plain way: there are no bridge axioms nor intricate interactions among modal operators. Therefore, soundness and completeness results are applicable as follows. Fix a finite number of agents to prevent possible infiniteness of the system. For the normal operators, apply the results in [1](see Appendix); for the logics of agency, apply [11]. The following theorem holds [6, Theorem 3]:

Theorem 1 (Temporalization/Modalization: Transfer of Complete Logics). *If* **N** *and Does are complete logics, so is* **N***(Does).*

Hence, **N**(Does) is complete, too.

4 Computing Collective Trust

Any possible computation model for collective trust requires that the underlying logic is at least decidable. In this section we exploit the following result [6, Theorem 4]:

Theorem 2 (Temporalization/Modalization: Transfer of Decidable Logics). *If* **N** *and Does are complete and decidable, so is* **N***(Does).*

Hence, we show that the logic **N**(Does) is also decidable by simply proving that the component logics (Does and **N**) are decidable. On account of this result, an algorithm for model checking is subsequently outlined.

4.1 Decidability

The logic for Does was proved in [11] to enjoy the finite model property and to be decidable. What about the logic **N**?

Also proving that **N** is decidable is not hard. Indeed, on account of [4], proving that **N** enjoys the finite model property trivially follows. Let us adjust the following definition introduced in [4]:

Definition 1 ([4]). *A set of formulas Σ closed for subformulas is closed if it satisfies the following properties:*

1. *if* $\mathrm{CBel}_G\phi \in \Sigma$ *then* $\mathrm{EBel}_G(\phi \wedge \mathrm{CBel}_G\phi) \in \Sigma$
2. *if* $\mathrm{EBel}_G\phi \in \Sigma$ *then* $\{\mathrm{Bel}_i\phi | i \in G\} \subseteq \Sigma$
3. *if* $\mathrm{MInt}_G\phi \in \Sigma$ *then* $\mathrm{EInt}_G(\phi \wedge \mathrm{MInt}_G\phi) \in \Sigma$
4. *if* $\mathrm{EInt}_G\phi \in \Sigma$ *then* $\{\mathrm{Int}_i\phi | i \in G\} \subseteq \Sigma$
5. *if* $\mathrm{JTrust}_y^G\phi \in \Sigma$ *then* $\{\mathrm{Trust}_y^i\phi | i \in G\}$
6. *if* $\mathrm{Rel}_y^G\phi \in \Sigma$ *then* $(\mathrm{JTrust}_y^G\phi \wedge \mathrm{MInt}_G(\mathrm{JTrust}_y^G\phi)) \in \Sigma$
7. *if* $\mathrm{CTrust}_s^G\phi in \Sigma$ *then* $(\mathrm{Rel}_s^G\phi \wedge \mathrm{CBel}_G(\mathrm{Rel}_s^G\phi)) \in \Sigma$.

Since we omit in **N** the operator Does, JTrust is defined here in terms of individual beliefs, intentions and goals. Rel is the mutual intention of JTrust, and CTrust is the common belief of Rel. On account of this simple observation, we can exactly proceed as done in [4] and establish the following result:

Lemma 1 ([4]). *Given a model*

$$\mathscr{M} = \langle W, \{B_i | i \in A\}, \{G_i | i \in A\}, \{I_i | i \in A\}, Val\rangle$$

let Σ *be a closed set of formulas and*

$$\mathscr{M}_\Sigma^f = \langle W^f, \{B_i^f | i \in A\}, \{G_i^f | i \in A\}, \{I_i^f | i \in A,\}, Val^f\rangle$$

be defined as follows:

- $W^f = W / \equiv_f^\Sigma$, $Val^f(a,[w]) = Val(a,w)$;
- $B_i^f = \{([w],[v]) \,|\, \forall \mathrm{Bel}_i\phi \in \Sigma, \mathscr{M}, w \models \mathrm{Bel}_i\phi \Rightarrow \mathscr{M}, v \models \phi, \forall X_i\phi \in \Sigma, \mathscr{M}, w \models X_i\phi \Leftrightarrow \mathscr{M}, v \models X_i\phi$ *where* $X \in \{\mathrm{Bel}, \mathrm{Goal}, \mathrm{Int}\}\}$;

- $G_i^f = \{([w],[v]) \mid \forall \text{Goal}_i\phi \in \Sigma, \mathscr{M}, w \models \text{Goal}_i\phi \Rightarrow \mathscr{M}, v \models \phi, \forall \text{Int}_i\phi \in \Sigma, \mathscr{M}, w \models \text{Int}_i\phi \Rightarrow \mathscr{M}, v \models \phi\}$;
- $I_i^f = \{([w],[v]) \mid \forall \text{Int}_i\phi \in \Sigma, \mathscr{M}, w \models \text{Int}_i\phi \Rightarrow \mathscr{M}, v \models \phi\}$.

The model $\mathscr{M}_\Sigma^f$ *thus defined is* a filtration of $\mathscr{M}$ through Σ.

From Lemma 1, it is an almost standard result to prove that the logic has the final model property and its satisfiability problem is decidable [4]. Due to the same reasons discussed in [4], also for each satisfiable formula ϕ of the logic **N** we can build a satisfying model of at most the size $O(2^{|\phi|})$, which however indicates that the following model checking algorithm has an exponential time complexity.

4.2 Model Checking

A model checker is a program that solves the model checking problem. The global model checking problem for **N**(Does) consists in checking whether, given a formula φ, and given $\mathfrak{M}$ model for **N**(Does), there exists a $w \in W$ such that $\mathfrak{M}, w \models \varphi$. We follow the modal model checker construction of [10]. Let φ be a formula and let $\text{MM}\mathscr{L}_{\text{Does}}(\varphi)$ be the set of *maximal monolithic subformulas of* φ belonging to $\mathscr{L}_{\text{Does}}$. Let φ' be the **N**-formula obtained by replacing every subformula $\alpha \in \text{MM}\mathscr{L}_{\text{Does}}(\varphi)$ by a new proposition letter p_α. Below are the sketches of the model-checkers needed to solve the modal checking problem for **N**(Does)[3]:

$Function\ MC_{\mathbf{N}(\text{Does})}((A, W, B_i, G_i, I_i, V', \{d_i\}), \varphi)$
  input: a modalized model $\mathfrak{M}$ and a formula $\varphi \in \mathscr{L}_{\mathbf{N}(\text{Does})}$
  compute $\text{MM}\mathscr{L}_{\text{Does}}(\varphi)$
  $for\ \ every\ \alpha \in \text{MM}\mathscr{L}_{\text{Does}}(\varphi)$
    $i := identify\ the\ agent\ involved\ in\ \alpha$
    $for\ \ every\ w \in W$
      $if(MC_{\text{Does}}(d_i(w), \alpha) = true)\ then$
        $V'(w) := V'(w) \bigcup \{p_\alpha\}$ */*fuzzling*/*

  build up φ' */* systematically replace variables generated above */*
  return $MC_{\mathbf{N}}((A, W, B_i, G_i, I_i, V', \{d_i\}), \varphi')$*;/*calls to the normal checker*/*

$Function\ MC_{\text{Does}}(d_i(w), \alpha)$
  input: a Scott-Montague model of structure η and
  a maximal monolithic sub-formula α.
  while there are neighbourhoods unchecked in $d_i(w)$
    $n_k = set\ n_i \in d_i(w)$ */**n_k *iterates on the set of neighbourhoods*/*
    $for\ every\ w \in n_k$
      $if\ \alpha \notin \mathrm{v}(w)\ then\ return\ false$
  return true

[3] To simplify the notation and have a more compact layout, we assume to work below in MC_{Does} with equivalent Scott-Montague models for Does and not with multi-relational ones. This assumption is non-problematic, since these semantics are equivalent.

$FunctionMC_{\mathbf{N}}((A,W,B_i,G_i,I_i,V',d_i),\varphi')$
  input: a model $\mathfrak{M} = (A,W,B_i,G_i,I_i,V',d_i)$ and a formula φ'
  $for\ \ every\ w \in W$
    $if\ check((A,w,B_i,G_i,I_i,V'),\varphi')$
      $return\ w$
  $return false$

$Function\ check((A,w,B_i,G_i,I_i,V'),\alpha)$
  $case\ on\ the\ form\ of\ \alpha$
    $\alpha = p_{\alpha'}:$
      $if\ p_{\alpha'} \notin V'(w)$
        $return\ false$
    $\alpha = \neg\alpha':$
      $if\ check((A,w,B_i,G_i,I_i,V'),\alpha')$
        $return\ false$
    $\alpha = \alpha_1 \wedge \alpha_2:$
      $if\ not\ check((A,w,B_i,G_i,I_i,V'),\alpha_1)\ or$
      $or\ not\ check((A,w,B_i,G_i,I_i,V'),\alpha_2)$
        $return\ false$
    $\alpha = \alpha_1 \vee \alpha_2:$
      $if\ not\ check((A,w,B_i,G_i,I_i,V'),\alpha_1)\ and$
      $and\ not\ check((A,w,B_i,G_i,I_i,V'),\alpha_2)$
        $return\ false$
    $\alpha = \mathrm{Bel}_i(\alpha'):$
      $for\ each\ v\ such\ that\ wB_iv$
        $if\ not\ check((A,v,B_i,G_i,I_i,V'),\alpha')$
          $return\ false$
    $\alpha = \mathrm{Goal}_i(\alpha'):$
      $for\ each\ v\ such\ that\ wG_iv$
        $if\ not\ check((A,v,B_i,G_i,I_i,V'),\alpha')$
          $return\ false$
    $\alpha = \mathrm{Int}_i(\alpha'):$
      $for\ each\ v\ such\ that\ wI_iv$
        $if\ not\ check((A,v,B_i,G_i,I_i,V'),\alpha')$
          $return\ false$
    $others:\ \ return\ false$
  $return\ true$

The procedures should be understood as follows. Given a modalized model and a formula φ, $MC_{\mathbf{N}(\mathrm{Does})}$ first computes the set $\mathrm{MM}\mathscr{L}_{\mathrm{Does}}(\varphi)$ of maximal monolithic subformulas of φ. For each of these, the checker identifies which agent is carrying out the action. Then, the checker establishes the worlds where that action has been carried out successfully. For doing this, the MC_{Does} checker is called with the Scott-Montague model $d_i(w)$ as parameter (recall d_i has structure η). MC_{Does} is nothing but pseudo-code for the valuation function v, it tests whether there is a neighborhood of w where α holds.

If so, the new letter p_α is added to $V'(w)$ to register such successful agency. Finally, before calling the normal model checker $MC_{\mathbf{N}}$, the new formula φ' is built without the Does modalities; these have been replaced in the former fuzzling.

5 Independent Combination of Mental States and Actions

$\text{Does}_i(\text{Goal}_j\ \mathscr{A})$ is a formula in which the normal modality appears within the scope of a non-normal Does. Note that, according to Observation 1, we cannot express this formula in the original system. An independent combination between a basic temporal and a simple deontic logic for MAS has been recently depicted in [20]. That combination puts together two normal modal logics: a temporal one and a deontic one.

Our aim now is to combine the normal and the non-normal counterparts of $\mathfrak{F}$ to get a new system where we can write and test the validity of formulas with arbitrarily interleaved cognitive and agency modalities.

For doing this, let us take a look to $\mathfrak{F}$ again. Consider it once more as a split into two separate substructures: one gathering the normal logics, and another one gathering the logics of agency. Again, there are two overlapping "nets" of relations identifiable over the same set W. The former is a Kripke-style cognitive ontology where goals, beliefs, intentions are interpreted, i.e., it captures internal (mental) motivational and informational aspects of agents (also the deontic aspects of the system, but recall that we do not explicitly consider them in this paper) the latter is a Scott-Montague structure which captures the external, visible, behavioral side of agents.

Now to the combination. First, duplicate and add subscripts to the elements in W to get one set of situations W_N, and another set W_D. Now build an ontology $W_N \times W_D$ of pairs (w_N, w_D).

Combination: Syntax Let $\mathscr{L}_{\mathbf{N}}$ denote the language of **N** (the base logic restricted to the normal operators), and $\mathscr{L}_{\text{Does}}$ denote the language of the logic of agency. The language $\mathscr{L}_{\mathbf{N}\times\text{Does}}$ is obtained by taking the union of the formation rules for the combination of $\mathscr{L}_{\mathbf{N}}$and $\mathscr{L}_{\text{Does}}$. Unlike the case of $\mathscr{L}_{\mathbf{N}(\text{Does})}$, $\text{Does}_i(\text{Goal}_j\ \mathscr{A})$ and $\text{Goal}_j(\text{Does}_i\ \mathscr{A})$ are both formulas of $\mathscr{L}_{\mathbf{N}\times\text{Does}}$.

Combination: Semantics Assume that we have two structures: $(A,\ W_N,\ \{B_i\},\ \{G_i\},\ \{I_i\},\ V',)$ and $(A, W_D, \{D_i\}, \mathtt{v})$, where to respectively test the validity of the normal modalities and the non-normal (Does) modalities. The former is a Kripke model; the latter a Scott-Montague model. Interpret $\mathscr{L}_{\mathbf{N}\times\text{Does}}$ formulas over a combined model

$$\mathfrak{C} = (A, W_N \times W_D, \{B_i\}_{i\in A}, \{G_i\}_{i\in A}, \{I_i\}_{i\in A}, \{D_i\}_{i\in A}, \mathtt{V}),$$

where:

- A is the set of agents;
- $W_N \times W_D$ is a set of pairs of situations;
- $\{B_i\}_{i\in A}, \{G_i\}_{i\in A}, \{I_i\}_{i\in A}$ are the accessibility relations for the normal operators (with semantics as in Section 3);
- $\{D_i\}_{i\in A}$ are the accessibility relations for the agency operators; and
- $\mathtt{V} : W_N \times W_D \to Pow(P)$ is a function assigning to each pair in $W_N \times W_D$ the set of proposition letters in P which are true.

The definition of a formula in $\mathscr{L}_{\mathbf{N}\times\mathrm{Does}}$ being satisfied in a model $\mathfrak{C}$ at state (w_N,w_D) amounts to:

$$\mathfrak{C},(w_N,w_D) \models \mathrm{Bel}_i\ \mathscr{A}\ \textit{iff}\ \forall v_N \in W_N(\textit{if } w_N B_i v_N \textit{ then } \mathfrak{C},(v_N,w_D) \models \mathscr{A}).$$
$$\mathfrak{C},(w_N,w_D) \models \mathrm{Goal}_i\ \mathscr{A}\ \textit{iff}\ \forall v_N \in W_N(\textit{if } w_N G_i v_N \textit{ then } \mathfrak{C},(v_N,w_D) \models \mathscr{A}).$$
$$\mathfrak{C},(w_N,w_D) \models \mathrm{Int}_i\ \mathscr{A}\ \textit{iff}\ \forall v_N \in W_N(\textit{if } w_N I_i v_N \textit{ then } \mathfrak{C},(v_N,w_D) \models \mathscr{A}).$$
$$\mathfrak{C},(w_N,w_D) \models \mathrm{Does}_i\ \mathscr{A}\ \textit{iff there exists a neighborhood } n \textit{ of } w_D \textit{ such that } \forall v \in n\ (\mathfrak{C},(w_N,v) \models \mathscr{A}).$$

A scan through the combined structure is done according to which operator is being tested. Normal operators move along the first component (w_N), and non-normal operators move along the second component of the current world (w_D).

Example 4 (Persuasion). The formula $\mathrm{Does}_i(\mathrm{Goal}_j\ \mathscr{A})$ can be seen as a form of persuasion, meaning that agent i makes agent j have $\mathscr{A}$ as goal. How do we test the validity of such a formula in a world (w_N,w_D)? The movements along the multi-graph are determined by $\mathfrak{C},(w_N,w_D) \models \mathrm{Does}_i\ (\mathrm{Goal}_j\ \mathscr{A})$ iff $\exists$ neighbourhood n_i of w_D such that $\forall v_k \in n_i\ (\mathfrak{C},(w_N,v_k) \models \mathrm{Goal}_j\ \mathscr{A})$, which amounts to test $\forall v_k \in n_i$ (iff $\forall u_N \in W_N$ (if $w_N G_j u_N$ then $\mathfrak{C},(u_N,v_k) \models \mathscr{A}$)).

6 Summary an Future Work

In this paper we have offered technical details for combining normal and a non-normal logics for modeling the notion of collective trust and for proving the completeness and decidability for the logic resulting form such a combination. Such combinations lead to different levels of expressiveness of the system by using temporalization and modalization techniques. On account of decidability results, we gave a possible structure for a combined model checker.

Let us consider three research issues for future work.

The Obl *modality.* We dealt with some of the modalities underlying the trust theory in [19]. In that work, a deontic connotation for the concept of collective trust is developed. Lawful support to collective trust is guaranteed in the theory with the schema: $(\mathrm{CTrust}_y^G\ \mathscr{A}) \to \mathrm{Obl}^G(\mathrm{Does}_y\ \mathscr{A})$, which is devised with a view to reflect the lawful force of trust, relativized to groups. The schema is to be understood as a standard of (good faith) behavior that can be identified with reference to social or group norms, to correctness, or reasonableness: if the group trusts agent y with respect to $\mathscr{A}$, agent y is obliged to carry out $\mathscr{A}$.

For capturing this deontic connotation of CTrust, we must consider deontic modalities such as Obl and Obl^G. Obl is the deontic operator for generic obligations, meaning “it is obligatory that” [18,14], and Obl^G is a relativized obligation operator which is meant to stand for “it is obligatory in the interest of G that” (see e.g. [13]). If these deontic modalities have the usual accepted KD and KD_n semantics, this extension is almost trivial: it is sufficient to add appropriate accessibility relations to the frames for **N**.

Things can get more complex if we characterize the deontic operators in weaker (non-normal) systems and apply combination techniques with more than just one non-normal modal logic [7].

Complexity. The proposed logic, though decidable, is EXPTIME complete. [4] proposed some methods for reducing this complexity, such as bounding modal depth of formulas and bounding the number of propositional atoms. It is an interesting research issue to check if these techniques can be useful also in the present framework.

Further combinations. Theoretically speaking, the very idea of reasoning about time should extend the current framework. For example, a basic temporalization amounts to place the temporal machinery on top of the modalized system, just in the same spirit we placed the normal machinery on top of the non-normal one. Consider the model $(T,<,g,t_0)$. The outer frame $(T,<)$ corresponds to the temporal evolution of the system; t_0 in T is the initial point in time. The system evolves through time in the sense that new groups and generic/individual beliefs, intentions, trust relations, obligations are settled while some others become obsolete. In its turn, g is the total function that brings in a model $\mathfrak{M}$ for each point in time.

A Completeness Proof for N

In this appendix, a completeness proof is sketched for the restriction **N**. The method used is often applied in modal logic for proving completeness with respect to finite models; is in turn inspired by the completeness proofs of mutual intentions shown by Dunin-Keplicz and Verbrugge in [3]. In fact, we adapt that to **N**, and apply Definition 4.24 and Theorem 4.22 described in [2]; these respectively settle how to construct a canonical model for a normal logic, and state that a normal modal logic is strongly complete with respect to its canonical model.

We have to prove that, supposing that $\mathbf{N} \nvdash \varphi$, there is a model $\mathfrak{M}_{\mathbf{N}}$ and a $w \in \mathfrak{M}_{\mathbf{N}}$ such that $\mathfrak{M}_{\mathbf{N}}, w \not\models \varphi$. The proof has four steps:

Step 1: Closure Construct a finite set of formulas Φ called the closure of φ. Φ contains φ and all its sub-formulas, plus certain other formulas that are needed in Step 4 below to show than an appropriate valuation falsifying φ at a certain world can be defined. The set Φ is also closed under single negations.

The closure of φ with respect to **N** is the minimal set Φ of **N**-*formulas* such that, for every agent, the following hold (see also Definition 1):

1. $\varphi \in \Phi$.
2. If $\psi \in \Phi$ and χ is a sub-formula of ψ, then $\chi \in \Phi$;
3. If $\psi \in \Phi$ and Φ itself is not a negation, then $\neg\psi \in \Psi$;
4. If $\mathrm{MInt}_G(\psi) \in \Phi$ then $\mathrm{EInt}_G(\psi \wedge \mathrm{MInt}_G(\psi)) \in \Phi$;
5. If $\mathrm{EInt}_G(\psi) \in \Phi$ then $\mathrm{Int}_i\psi \in \Phi$ for all $i \in G$;

6. $\neg\mathrm{Int}_i \perp \in \Phi$ for all $i \leq m$;
7. If $\mathrm{CBel}_G(\psi) \in \Phi$ then $\mathrm{EBel}_G(\psi \wedge \mathrm{CBel}_G(\psi)) \in \Phi$;
8. If $\mathrm{EBel}_G(\psi) \in \Phi$ then $\mathrm{Bel}_i, \psi \in \Phi$ for all $i \in G$;
9. $\neg\mathrm{Bel}_i \perp \in \Phi$ for all $i \leq m$;
10. $\neg\mathrm{Goal}_i \perp \in \Phi$ for all $i \leq m$.

It should be clean that for every formula ϕ, Φ is a *finite* set of formulas (recall that the language in [19] includes: MInt, EInt, EBel).

Step 2: Canonical model. To construct a canonical model we need to define the worlds and relations between them. Each of these worlds are *maximally* **N**-*consistent* sets. To build this sets, we apply the Lindenbaum Lemma (which is proved in Lemma 4.17 [2]) over Φ *step 1*, as follows:

Let Φ be the closure of ϕ with respect to **N**. If $\Gamma \subseteq \Phi$ is **N**-*consistent*, then there is a set $\Gamma' \supseteq \Gamma$ which is maximally **N**-*consistent* in Φ.

Step 3: Build a canonical model using Definition 4.24 [2]. . This model will turn out to contain a world where $\neg\psi$ holds. Let $\mathfrak{M}_\varphi =< S_\phi, \pi, I_1, ..., I_m, B_1, ..., B_m, G_1, ..., G_m >$ be the Kripke model defined as follows:

- As domain of states, one state s_Γ is defined for each *maximally* **N**-*consistent* $\Gamma \subseteq \Phi$. Note that, because Φ is finite, there are only finitely many states. Formally, we defined $CON_\Phi = \{\Gamma | \Gamma$ is *maximally* **N**-*consistent* in $\Phi\}$ and $S_\varphi = \{s_\Gamma | \Gamma \in CON_\Phi\}$.
- To make a truth assignment π, we want to conform to the propositional atoms that are contained in the maximally consistent sets corresponding to each world. Thus we define $\pi(s_\Gamma)(p) = 1$ if and only if $p \in \Gamma$. Note that this makes all propositional atoms that do not occur in φ false in every world of the model.
- The corresponding relations are defined as follows:

$$I_i = \{(s_\Gamma, s_\triangle) | \psi \in \triangle \ for\ all\ \psi\ such\ that\ \mathrm{Int}_i(\psi) \in \Gamma\}$$
$$B_i = \{(s_\Gamma, s_\triangle) | \psi \in \triangle \ for\ all\ \psi\ such\ that\ \mathrm{Bel}_i(\psi) \in \Gamma\}$$
$$G_i = \{(s_\Gamma, s_\triangle) | \psi \in \triangle \ for\ all\ \psi\ such\ that\ \mathrm{Goal}_i(\psi) \in \Gamma\}$$

It will turn out that with this definition we get $\mathfrak{M}_\varphi, s_\Gamma \models p$ iff $p \in \Gamma$ for propositional atoms p.

Step 4: Completeness of **N**. If $\mathbf{N} \nvdash \varphi$ then there is a model $\mathfrak{M}$ and a w such that $\mathfrak{M}, w \not\models \varphi$. Proof: Suppose $\mathbf{N} \nvdash \varphi$. Take $\mathfrak{M}_\varphi$ as in step 3. Note that there is a formula χ logically equivalent to $\neg\varphi$ that is an element of Φ; if ψ does not start with a negation, χ is the formula $\neg\varphi$ itself. Now, using the Lindenbaum Lemma, there is a maximally consistent set $\Gamma \subseteq \Phi$ such that $\chi \in \Gamma$. By the Finite Truth Lemma, if $\Gamma \in CON_\phi$ then for all $\psi \in \Phi$ it holds that $\mathfrak{M}_\varphi, s_\Gamma \models \psi$ iff $\psi \in \Gamma$. Thus, this implies that $\mathfrak{M}_\varphi, s_\Gamma \models \chi$, thus $\mathfrak{M}_\varphi, s_\Gamma \not\models \varphi$. Details of the Finite Truth Lemma proof are left to the reader (see [3] and [2], Lemma 4.21).

References

1. Ambrossio, A., Mendoza, L.: Completitud e implementación de modalidades en MAS. Thesis report, Facultad de Informatica UNLP (2011)
2. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science, vol. 53. Cambridge University Press, Cambridge (2001)
3. Dunin-Keplicz, B., Verbrugge, R.: Collective intentions. Fundam. Inform., 271–295 (2002)
4. Dziubinski, M., Verbrugge, R., Dunin-Keplicz, B.: Complexity issues in multiagent logics. Fundam. Inform. 75(1-4), 239–262 (2007)
5. Elgesem, D.: The modal logic of agency. Nordic Journal of Philosophical Logic 2, 1–46 (1997)
6. Fajardo, R., Finger, M.: Non-normal modalisation. In: Advances in Modal Logic 2002, pp. 83–96 (2002)
7. Fajardo, R., Finger, M.: How not to combine modal logics. In: IICAI, pp. 1629–1647 (2005)
8. Falcone, R., Castelfranchi, C.: Social Trust: A Cognitive Approach. In: Trust and Deception in Virtual Societies, pp. 55–90. Kluwer Academic Publishers (2001)
9. Finger, M., Gabbay, D.: Combining temporal logic systems. Notre Dame Journal of Formal Logic 37 (1996)
10. Franceschet, M., Montanari, A., De Rijke, M.: Model checking for combined logics with an application to mobile systems. Automated Software Eng. 11, 289–321 (2004)
11. Governatori, G., Rotolo, A.: On the Axiomatization of Elgesem's Logic of Agency and Ability. Journal of Philosophical Logic 34(4), 403–431 (2005)
12. Hansson, B., Gärdenfors, P.: A guide to intensional semantics. Modality, Morality and Other Problems of Sense and Nonsense. Essays Dedicated to Sören Halldén (1973)
13. Herrestad, H., Krogh, C.: Deontic logic relativised to bearers and counterparties. In: Bing, J., Torvund, O. (eds.) Anniversary Antology in Computers and Law, TANO (1995)
14. Jones, A., Sergot, M.: A logical framework. In: Open Agent Societies: Normative Specifications in Multi-agent Systems (2007)
15. Memmo, D., Sartor, G., di Cardano, G.Q.: Trust, Reliance, Good Faith, and the Law. In: Nixon, P., Terzis, S. (eds.) iTrust 2003. LNCS, vol. 2692, pp. 150–164. Springer, Heidelberg (2003)
16. Rasga, J., Sernadas, A., Sernadas, C.: Fibring as biporting subsumes asymmetric combinations. Preprint, SQIG - IT and IST - TU Lisbon, 1049-001 Lisboa, Portugal (2010)
17. Rasga, J., Sernadas, A., Sernadas, C.: Importing logics. Studia Logica (2011)
18. Rotolo, A., Sartor, G., Smith, C.: Good faith in contract negotiation and performance. IJBPIM Journal of Business Processes and Management, Special Issue on Contract Architectures and Languages 4, 154–173 (2009)
19. Smith, C., Rotolo, A.: Collective trust and normative agents. Logic Journal of IGPL 18(1), 195–213 (2010)
20. Smith, C., Rotolo, A., Sartor, G.: Representations of time within normative MAS. Frontiers in Artificial Intelligence and Applications 223, 107–116 (2010)
21. Whittaker, S., Zimmermann, R.: Good Faith in European Contract Law. Cambridge University Press, Cambridge (2000)