A Normal Modal Logic for Trust in the Sincerity

Christopher Leturc
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
Caen, France
christopher.leturc@unicaen.fr

Grégory Bonnet
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
Caen, France
gregory.bonnet@unicaen.fr

ABSTRACT

In the field of multi-agent systems, as some agents may be not reliable or honest, a particular attention is paid to the notion of trust. There are two main approaches for trust: trust assessment and trust reasoning. Trust assessment is often realized with fuzzy logic and reputation systems which aggregate testimonies – individual agents' assessments – to evaluate the agents' global reliability. In the domain of trust reasoning, a large set of works focus also on trust in the reliability as for instance Liau's BIT modal logic where trusting a statement means the truster can believe it. However, very few works focus on trust in the sincerity of a statement – meaning the truster can believe the trustee believes it. Consequently, we propose in this article a modal logic to reason about an agent's trust in the sincerity towards a statement formulated by another agent. We firstly introduce a new modality of trust in the sincerity and then we prove that our system is sound and complete. Finally, we extend our notion of individual trust about the sincerity to shared trust and we show that it behaves like a KD system.

KEYWORDS

Logics for agents and multi-agent systems; Trust and reputation.

ACM Reference Format:
Christopher Leturc and Grégory Bonnet. 2018. A Normal Modal Logic for Trust in the Sincerity. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

In the field of multi-agent systems, a particular attention was paid to the notion of trust. Indeed, in many multi-agent systems, agents must cooperate with each other in order to satisfy their goals. However, not all agents are necessarily reliable or cooperative and one of the main technique for determining whether an agent is reliable or not is to use a reputation system [20]. In such systems, agents which interact evaluate each other with a *trust* value which is refined as new interactions happen. Agents can then exchange those values with *testimonies*: a communication in which an agent tells if it trusts another agent. The aggregation of those testimonies provides a *reputation* value, which represents a notion of collective trust. Generally, an agent with a high reputation can be trusted by other agents, even if those latter has never interacted with the former before. It is important to notice that, as reputation systems aim at approximating the reliability of the agents, testimonies do not represent a ground truth: they represent a subjective evaluation of a single interaction. Consequently, taking into account different (even contradictory) testimonies about an agent is of interest. Moreover, as testimonies are subjective evaluations, they can be biased by the agents' capabilities or intentions (for manipulation purpose for instance). To mitigate this problem, classical reputation systems weigh the testimonies with respect to the agent's reputation [11, 12, 19]. However, several works show the interest to clearly differentiate trust in reliability and trust in honesty of a testimony, this latter being called *credibility* [14, 18, 21, 24, 26].

In the literature dedicated to model the socio-cognitive aspects of trust [1], some works are interested to model how to assess trust [6, 8, 13, 25]. Some other works focus on reasoning about trust instead. They are interested to model trust with modal logics [7, 10, 22, 23] and to characterize what are the logical implications to trust another agent. Those modal logics make it possible to express trust by means of one or more modalities such as intentions, beliefs, goals or acts. While those approaches make it easy to express some aspects of trust such as *delegation*, they focus on trusting the actions of other agents. However, in reputation systems, agents are required to communicate testimonies to inform the other agents, for example, of the quality of the services offered by third-party agents. To address this problem, some works are devoted to knowledge revision based on trust [9, 16], and others are devoted to modeling the trust an agent expresses about the discourse of another agent [3–5, 15].

While most of those latter deal with trust in the reliability of an agent when he communicates a proposition, very few works deal with trust in the honesty. For instance, Liau [15] defines a trust modality to express trust in the judgment of another agent over a proposition. By this modality, he understands the trust granted by an agent to the reliability of a discourse of another agent, which is indeed well different from the trust granted to the honesty of an agent. Thus, based on the Oxford dictionary definition of honesty – "free of deceit, truthful and sincere" –, we propose a first step with a modal logic expressing the *trust in the sincerity* granted by an agent i to a statement ϕ proposed by another agent j. The main characteristic of our logic is to link trust modality with the beliefs of the trusted agent: an agent is sincere if it believes what it says. We prove the soundness and completeness of our logic, along with several interesting properties, such as the non-transitivity of trust in the sincerity: it is not because an agent i trusts in the sincerity of an agent j about its trusts in the sincerity of an agent k that the agent i should trust in the sincerity of the agent k.

This article is structured as follows. In Section 2, we present a state-of-the-art on modal logics for trust then, in Section 3, we present the semantics and the axiomatics associated with our logic. We show that our semantics is sound and complete in Section 4 and we give several properties of our logic in Section 5.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

2 MODAL LOGICS OF TRUST

Castelfranchi and Falcone [1] studied various fundamental components of trust, including its dynamic aspects, particularly in the context of decision-making and the construction of intentions. They also studied the act of self-trust, or how to authorize the delegation of shared tasks. They show that modeling trust in its entirety is very complex. Thus, most logical approaches restricted themselves to specific aspects of trust. We distinguish in this section two types of approaches: modal logics that use predicates to represent trust and modal logics that rely on a trust modality.

2.1 Trust as a predicate

Herzig *et al.* [10] consider trust as a predicate, meaning that agent i trusts another agent j about an action α having for consequences the proposition ϕ if, and only if, all the following statements are true:

(1) i has the goal ϕ,
(2) i believes that:
  (a) j is able to execute the action α,
  (b) j by doing α will ensure ϕ,
  (c) j intends α.

This makes it possible to define a predicate of *occurrent trust*. This notion reflects an aspect of trust in the present, namely the fact that an agent j is well-prepared to perform the action for which i trusts it. A second notion of trust, *dispositional trust*, expresses the trust granted by an agent i to an agent j about the fact that this agent j will realize the proposition ϕ in a specific context. Smith *et al.* [23] also consider an occurrent notion of trust meaning that an agent i trusts another agent j for ϕ if, and only if, all the following statements are true:

(1) i has the goal ϕ,
(2) i believes that j performs ϕ,
(3) i intends that:
  (a) j performs ϕ,
  (b) i does not perform ϕ.
(4) i has the goal that j intends ϕ,
(5) i believes that j intends ϕ.

Let us remark both trust notions are *trust in the reliability* of an agent. Moreover, those trusts cannot characterize trust in the reliability of a *communication*. Indeed, the previous definitions assume that an agent j is communicating ψ to an agent i, denoted $\alpha_{j,i}$ then the communication's consequences are $\phi = B_i\psi$. By applying the definition of Herzig *et al.*, i trusting j for the action $\alpha_{j,i}$ implies that i aims to believe the proposition ψ. Thus, it expresses the fact that the agent trusts because it has to believe as a goal.

Some other works, such as those of Christianson and Harbison [3] or Demolombe [5], propose to model other aspects of trust. Interestingly, Demolombe proposes a model for trust in the sincerity and trust in the honesty. His logic relies on several modalities – K_i, B_i, $Com_{i,j}$, O, P, and E_i which are respectively Knowledge, Belief, Communication, Obligation, Permission, and Bringing it about – and defines predicates for trust. On a first hand, *trust in the honesty* is defined as:

$$Thon_{i,j}(\phi) \triangleq K_i(E_j\phi \Rightarrow PE_j\phi)$$

It means an agent i trusts in the honesty of j if, and only if, i knows that if j brings it about that ϕ then j is allowed to bring it about that ϕ. Here, trust in the honesty is related to the noninfringement of norms, and does not encompass all aspects of honesty. On the other hand, *trust in the sincerity* is defined as:

$$Tsinc_{i,j}(\phi) \triangleq K_i(Com_{j,i}\phi \Rightarrow B_j\phi)$$

It means an agent i trusts j when i knows that if j communicates ϕ to j, then j believes ϕ. Although, it captures the notion of sincerity, the predicate is linked to a communication action modality associated with a minimal semantics. Consequently, it makes the trust predicate dependent of the communication axiomatic and $Tsinc$ cannot behaves like a KD system, which is important to not trust in the sincerity of an agent if it says something and its contrary.

2.2 Trust as a modality

Expressing trust of an agent i towards an agent j with a modality allows for expressing inference mechanisms that are necessary when we consider a trust aspect like a *disposition of an agent to act* [22] or a *reliability of information* [4, 7, 15].

The first approach, proposed by Singh [22], expresses a dispositional trust through a modality $T^d_{i,j}(\phi, \psi)$ meaning that an agent i trusts another agent j to realize ψ in a context ϕ. If ϕ is true, the trust of agent i towards j is activated. An occurrent trust can be expressed then as $T^d_{i,j}(\top, \psi)$ meaning that at every moment (and therefore in the present moment) i trusts j about the statement ψ. Singh's approach uses around twenty axioms. For example, if ψ is already true then the agent i does not trust j so that, in the context ϕ, ψ is true.

The second approach, proposed by Liau [15] and extended by Dastani *et al.* [4], introduces the BIT formalism for reasoning about trust of an agent i in the judgment of another agent j. This modality $T^r_{i,j}$ is associated with a minimal semantics as the trust may be irrational: an agent can trust another agent that says something and its contrary ($T^r_{i,j}p \wedge T^r_{i,j}\neg p$ is not inconsistent). In order to deduce new beliefs thanks to information acquisition and trust, Liau introduces a modality $I_{i,j}$ which means that i has acquired information from j. While Demolombe uses a minimal semantics for $Com_{i,j}$, Liau defines $I_{i,j}$ as a KD system representing the consequences of a successful communication. Interestingly, BIT is extended with several axiomatic systems such as BA, TR, SY, in order to catch specific aspects of trust in the reliability. For instance, BA is the less restrictive system: it considers one axiom for trust to infer new beliefs, i.e. $\vdash_{BA} B_iI_{i,j}\phi \wedge T^r_{i,j}\phi \Rightarrow B_i\phi$, and one other axiom to represent self-awareness of the granted trust, i.e. $\vdash_{BA} T^r_{i,j}\phi \equiv B_iT^r_{i,j}\phi$. As another example, the SY system captures the case where an agent can trust the reliability of another agent for both ϕ and $\neg\phi$ in order to acquire new knowledge when asking a question: $\vdash_{SY} T^r_{i,j}p \Rightarrow T^r_{i,j}\neg p$. This axiom is highly relevant in reputation systems as it allows to acquire new knowledge without knowing in advance the response given by the agent. Indeed, when an agent i questions an agent j about a proposition p (denoted that $Q_{i,j}p$), Liau asserts that if agent i trusts j for its ability to answer the question, that is, if i trusts the judgment of j for p then i also trusts the judgment of j for $\neg p$.

Even the BIT formalism clearly takes the perspective of acquiring new information, its trust modality cannot express trust for two

agents that say propositions which contradict each other. In this sense, this approach does not deal with trust in the sincerity but trust in the reliability. Hence we propose the opposite: being able to trust agents that contradict each other and being unable to trust an agent that contradicts itself. This notion makes sense when agents are aware that they can be deceived by other agents. Indeed, we consider that, since trust is often accompanied by a risk, it is necessary to have rules of inference preventing the agents from blindly trusting another agent which contradicts itself from the sincerity perspective. However, it should be noted that trust in the reliability and trust in the sincerity are related.

Thus, we propose a *normal multi-modal logic* with a modality that allows representing trust in the sincerity of a discourse produced by an agent. More precisely, we want a truster agent being able to trust several agents that may contradict each other, as the sincerity is not related to truth: an agent may be wrong while being sincere.

3 A NORMAL MODAL LOGIC OF TRUST

In our modal logic, trust between agents is expressed by a modality $T^s_{i,j}\phi$ which means that i trusts in the sincerity of j about *a proposition* ϕ. As we also consider a belief modality B_i, we call our system TB (trust and belief).

3.1 Language

Let us define a language $\mathcal{L}_{T,B}$ which considers a set of propositional letters $\mathcal{P} = \{a, b, c, ...\}$, a set of agents $\mathcal{N}$ with $i, j \in \mathcal{N}$ two agents, and $p \in \mathcal{P}$ a propositional variable. We consider the following BNF grammar rule :

$$\psi ::= p|\neg\psi|\psi \wedge \psi|\psi \vee \psi|\psi \Rightarrow \psi|T^s_{i,j}\psi|B_i\psi$$

Let us notice that B_i differs from $T^s_{i,i}$ because the latter means that an agent i trusts itself in its sincerity about ϕ. Furthermore, unlike Liau and Demolombe [5], we do not introduce explicitly an information acquisition or communication modality. Firstly, Liau considers trust as a potential trust modality, meaning that if an agent trusts in the judgement of another agent then the truster can (i.e. has the potential to) believe the trustee for its answer. Our trust modality $T^s_{i,j}$ is different as it is an effective trust modality which is active in the present time: when it is the case that an agent trusts in the sincerity of another agent for ϕ, it means that the truster believes that the trustee believes what it said. Secondly, as we focus on logical implications of trust, we define an atomic fragment of a modal logic for trust in the sincerity. We consider that when an agent i trusts in the sincerity of another agent j about ϕ, i has already acquired information from j to deduce if j is sincere or not. Therefore, we do not need a specific modality to represent information acquisition.

3.2 Associated Kripke semantics

We define a Kripke frame $C = (\mathcal{W}, \{\mathcal{B}_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}})$ associated with $\mathcal{L}_{T,B}$ where:

- $\mathcal{W}$ is a non-empty set of possible worlds,
- $\{\mathcal{B}_i\}_{i\in\mathcal{N}}$ is a set of binary relations such that:

$$\forall i \in \mathcal{N}, \forall w \in \mathcal{W} : \mathcal{B}_i(w) := \{v \in \mathcal{W} | w\mathcal{B}_i v\}$$

- $\{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}}$ is a set of binary relations such that:

$$\forall i, j \in \mathcal{N}, \forall w \in \mathcal{W} : \mathcal{T}^s_{i,j}(w) := \{v \in \mathcal{W} | w\mathcal{T}^s_{i,j} v\}$$

We define a Kripke model as $\mathcal{M} = (\mathcal{W}, \{\mathcal{B}_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}}, i)$ with $i : \mathcal{P} \to 2^{\mathcal{W}}$ an interpretation function. For each world $w \in \mathcal{W}$, for all $\phi, \psi \in \mathcal{L}_{T,B}$ and for all $p \in \mathcal{P}$:

(1) $w \models \top$
(2) $w \not\models \bot$
(3) $w \models p$ iff $w \in i(p)$
(4) $w \models \neg\phi$ iff $w \not\models \phi$
(5) $w \models \phi \vee \psi$ iff $w \models \phi$ or $w \models \psi$
(6) $w \models \phi \wedge \psi$ iff $w \models \phi$ and $w \models \psi$
(7) $w \models \phi \Rightarrow \psi$ iff $w \models \neg\phi$ or $w \models \psi$
(8) $w \models B_i\phi$ iff $\forall v \in \mathcal{W} : w\mathcal{B}_i v, v \models \phi$
(9) $w \models T^s_{i,j}\phi$ iff $\forall v \in \mathcal{W} : w\mathcal{T}^s_{i,j} v, v \models \phi$

Let us notice that B_i is a classical $\Box$ modality like [5, 10, 15]. Concerning the trust modality, we consider an accessibility relation for each pair of agents $(i, j) \in \mathcal{N}^2$. This binary relation translates that an agent i trusts j for a property ϕ in a possible world $w \in \mathcal{W}$ if, and only if ϕ is true in each accessible world from w by the relation $\mathcal{T}^s_{i,j}$.

Classicaly, ϕ is satisfiable in w iff $w \models \phi$ is true, and ϕ is valid in a model $\mathcal{M}$ (written $\mathcal{M} \models \phi$) iff, for each world $w \in \mathcal{W}$, $\mathcal{M}, w \models \phi$. A formula ϕ is valid in a frame C (written $\models_C \phi$ or $C \models \phi$) iff, for each model $\mathcal{M}$ based on the frame C, $\mathcal{M} \models \phi$.

Our Kripke frame C is such that for all $i, j \in \mathcal{N}$:

(1) $\forall w \in \mathcal{W}, \exists v \in \mathcal{W} : w\mathcal{T}^s_{i,j} v$
(2) $\forall w, u, v \in \mathcal{W} : w\mathcal{B}_i u \wedge u\mathcal{T}^s_{i,j} v \Rightarrow w\mathcal{T}^s_{i,j} v$
(3) $\forall w, u, v \in \mathcal{W} : w\mathcal{B}_i u \wedge w\mathcal{T}^s_{i,j} v \Rightarrow u\mathcal{T}^s_{i,j} v$
(4) $\forall w, u, v \in \mathcal{W} : w\mathcal{B}_i u \wedge u\mathcal{B}_j v \Rightarrow w\mathcal{T}^s_{i,j} v$
(5) $\mathcal{B}_i$ is serial, transitive and Euclidean.

When an agent trusts in the sincerity of another agent, it takes the risk of being deceived. Thus, a way to be protected from deception is to not be able to trust in something and its opposite. Indeed, an agent cannot trust another one which contradicts itself. A glaring example of this connection between trust in the sincerity and non-contradiction is very well illustrated by a police investigation into a crime scene. The police officers trust in the sincerity of the witnesses as long as they do not get contradictory information. Therefore, a way to consider this principle is to say that there is always an accessible world by $\mathcal{T}^s_{i,j}$ from any world, which is given by property (1).

An agent is also aware of the trust it grants to another agent. The property (2) given in [7, 15] illustrates this constraint: if an agent is trusted then the truster agent believes that it trusts the trustee. Moreover, we add the property (3) which means that if an agent does not trust another agent then the former agent believes that it does not trust the latter agent.

The property (4) is associated with the notion of sincerity underlying in honesty: a sincere agent communicates information it believes true. Thus, when an agent trusts another one for ϕ then it can deduce that it believes the other agent believes ϕ.

Finally, the last properties given in (5) are the usual properties used to represent a doxastic modality [5, 10, 15].

3.3 Axiomatic system

We consider the following axioms: propositional calculus tautologies, classical rules of inference in modal logics (**K**, **Nec**, **Sub**) and a consistency axiom between trusts (**D**). Our logic of trust is therefore a normal logic that satisfies the necessitation, substitution, modus ponens and the Kripke's axiom K. The **necessitation** means that if a formula ϕ is a theorem ($\vdash \phi$) then any agent i can trust any other agent j about this theorem ($\vdash T^s_{i,j}\phi$) and any agent i believes ϕ ($\vdash B_i\phi$) . The **substitution** means that if we uniformly substitute any formula for any propositional letter in a theorem, the resulting formula is also a theorem. The **modus ponens** means that if a proposition $\vdash \phi$ is proved as a theorem and if it is also proved that $\vdash \phi \Rightarrow \psi$ is a theorem then the formula $\vdash \psi$ is proved. Furthermore we consider the definition of the derivation proof from the sets of formulas with necessitation(s) restricted to theorems in order that the deduction theorem follows easily.

Finally, our trust modality satisfies the axiom **K**: if an agent i trusts an agent j on p:= "The financial situation of company X is excellent." which implies q:="It is worthwhile to invest in company X." then, if i trusts j for p then i also trusts j for q. Formally,

$$\vdash T^s_{i,j}(p \Rightarrow q) \Rightarrow T^s_{i,j}p \Rightarrow T^s_{i,j}q \quad (K)$$

Let us notice that Liau [15] does not consider the axiom K. Instead, he uses a minimal semantics. Indeed, according to Liau, when considering $T^r_{i,j}p \wedge T^r_{i,j}(p \Rightarrow q)$, $T^r_{i,j}q$ must not be deduced: it is not because i trusts the judgment of j for both propositions p and $(p \Rightarrow q)$ that i trusts j for q. Considering artificial systems, an agent j that does not deduce ψ is an irrational agent. However, we can reasonably assume that all agents, in an artificial agent system, are rational. Thus, if i trusts in the judgement of j for both p and $p \Rightarrow q$ then i should trust j for q. In the context of sincerity, the same argument holds.

3.4 Non-inconsistency of trust

We want to express the fact that if an agent i trusts j for a proposition, i cannot trust j for the opposite because of reasons of coherence of the discourse: it is not possible to trust in the sincerity of an agent that contradicts itself.

$$\vdash T^s_{i,j}p \Rightarrow \neg T^s_{i,j}\neg p \quad (D)$$

This translates the fact that for instance if an agent i trusts in the sincerity of j for p :="The work is done" then this agent i does not trust in the sincerity of j for $\neg p$: a sincere agent must have a consistent discourse. However, we cannot generalize this axiom to any other agent $k \in \mathcal{N}$, $T^s_{i,j}p \nRightarrow \neg T^s_{i,k}\neg p$. Indeed, if an agent i trusts in the sincerity of j for p is true, that is $T^s_{i,j}p$. Nothing tells us and prevents us from having $T^s_{i,k}\neg p$ for another agent k and it is not an inconsistency situation. Indeed, since $T^s_{i,j}$ is trust in the sincerity, two agents may have contradictory discourses and it does not mean that they are not sincere. Moreover, if we assume such a generalization, we would immediately deduce the theorem $T^s_{i,j}p \wedge T^s_{i,k}\neg p \Rightarrow \neg T^s_{i,k}\neg p \wedge T^s_{i,k}\neg p$ which is generally not true.

Let us recall that Liau's BIT system does not allow to trust two different and contradictory sources, whereas in our case it is quite possible to trust them. On the contrary, Liau's model can trust an agent for something and its contrary whereas we cannot.

3.5 Link between trust and belief

Liau [15] has axiomatized a link between trust and belief. An agent is self-aware about the trust it grants to another agent. For instance, we consider that if an agent i trusts in the sincerity of j about the proposition p :="The product is good", then the agent i believes that i trusts in the sincerity of j on the proposition p:

$$\vdash T^s_{i,j}p \Rightarrow B_i T^s_{i,j}p \quad (4_{T,B})$$

However, instead of considering the reciprocal as Liau does, we consider a kind of *negative introspection*.

$$\vdash \neg T^s_{i,j}p \Rightarrow B_i \neg T^s_{i,j}p \quad (5_{T,B})$$

Interestingly, we show in Section 5 that our system allows us to deduce the reciprocals of both previous axioms.

Note that we do not consider an axiom of non-inconsistency between trust and belief. In fact, if an agent believes that something is true, it does not imply that it does not trust another agent that announces the opposite of his belief, i.e. $\forall i, j \in \mathcal{N}, B_i p \nRightarrow \neg T^s_{i,j}\neg p$ is not true in general. Indeed, in trust in the sincerity, an agent can believe p :="He is a good mechanic" and can trust in the sincerity of another agent for its opposite $\neg p$ at the same time.

Finally, a last important axiom is the *axiom of sincerity* associated with our modality of trust in the sincerity. It expresses the fact that if an agent i trusts in the sincerity of another agent j for p :="He told me the truth" then i believes that j believes that p.

$$\vdash T^s_{i,j}p \Rightarrow B_i B_j p \quad (S)$$

We do not consider the reciprocal of the axiom of sincerity. Let us recall the axiom deals with trust in the sincerity and not in the sincerity in itself. As trust is a special kind of mental state, between knowledge and belief where trust is weaker than knowledge, and belief is weaker than trust, it is possible for external reasons that an agent is wrong about its beliefs about the other agents. As it knows that its beliefs are not necessarily true, then the agent is free to not trust the others in order to protect itself.

Furthermore this property cannot be expressed in Liau's system. Indeed, even if it is not contradictory to write $T^r_{i,j}p \wedge T^r_{i,j}\neg p$, by considering $T^r_{i,j}p \Rightarrow B_i B_j p$ we would have $(T^r_{i,j}p \wedge T^r_{i,j}\neg p) \Rightarrow (B_i B_j p \wedge B_i B_j \neg p)$ which may be not true and so cannot be a theorem. In our model, it may be considered as a theorem, because even if $B_i B_j p \wedge B_i B_j \neg p$ is a contradiction, the false implies the false is always verified. In the same way, we are able to deduce the following theorem $\vdash T^s_{i,j}p \wedge T^s_{i,k}\neg p \Rightarrow (B_i B_j p \wedge B_i B_k \neg p)$ which is not a theorem in Liau's BA system [15] because, in this model, agents cannot trust the reliability of two different inconsistent sources ($\vdash_{BA} B_i(I_{i,j}\phi \wedge I_{i,k}\neg p) \Rightarrow \neg(T^r_{i,j}p \wedge T^r_{i,k}\neg p)$).

4 SOUNDNESS AND COMPLETENESS

Firstly, we prove the main validity results for our TB system, and recall the standard validity properties, characterizing the properties that must respect the accessibility relationships of our Kripke frame. Then we prove that our axiomatic system TB is sound. Finally, we demonstrate that the properties of those relationships completely describe the axiomatic system we proposed.

4.1 Valid formulas in our Kripke frame

We consider a frame $C = (\mathcal{W}, \{\mathcal{B}_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}})$ on $\mathcal{L}_{T,B}$. Let us first prove the corresponding Kripke frame for the axiom $(5_{T,B})$ $\vdash \neg T^s_{i,j}p \Rightarrow B_i\neg T^s_{i,j}p$.

Proposition 4.1. *For all agents $i, j \in \mathcal{N}, C \models \neg T^s_{i,j}p \Rightarrow B_i\neg T^s_{i,j}p$ if, and only if:*

$$\forall w, u, v \in \mathcal{W}, w\mathcal{B}_i u \wedge w\mathcal{T}^s_{i,j}v \Rightarrow u\mathcal{T}^s_{i,j}v$$

Proof. Let $i, j \in \mathcal{N}$ be two agents,

($\Rightarrow$) By contraposition, let us consider there exists $w, u, v \in \mathcal{W}$: $w\mathcal{B}_i u \wedge w\mathcal{T}^s_{i,j}v \wedge \neg(u\mathcal{T}^s_{i,j}v)$. Let us define a model $\mathcal{M}$ where $i(p) = \mathcal{W} \setminus \{v\}$. Since $i(p) = \mathcal{W} \setminus \{v\}$ and $w\mathcal{T}^s_{i,j}v$ we have $\mathcal{M}, w \models \neg T_{i,j}p$. Furthermore $\neg(u\mathcal{T}^s_{i,j}v)$ and, thus $\mathcal{M}, u \models T_{i,j}p$. Then since $w\mathcal{B}_i u$, we deduce $\mathcal{M}, w \models \neg B_i\neg T^s_{i,j}p$.

Consequently, there exists a model $\mathcal{M}$ and a world $w \in \mathcal{W}$ such that $\mathcal{M}, w \models \neg T^s_{i,j}p \wedge \neg B_i\neg T^s_{i,j}p$ i.e. $C \not\models \neg T^s_{i,j}p \Rightarrow B_i\neg T^s_{i,j}p$

($\Leftarrow$) By contraposition, there exists a model $\mathcal{M} = (\mathcal{W}, \{\mathcal{B}_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}}, i)$ and a world $w \in \mathcal{W}$ such that $\mathcal{M}, w \models \neg T^s_{i,j}p \wedge \neg B_i\neg T^s_{i,j}p$. Thus, there exists $v \in \mathcal{W}$, $w\mathcal{T}^s_{i,j}v$ such that $\mathcal{M}, v \models \neg p$ and there exists $u \in \mathcal{W}$: $w\mathcal{B}_i u$ such that $\mathcal{M}, u \models T^s_{i,j}p$. Since $v \notin i(p)$ and $\forall u' \in \mathcal{W}$: $u\mathcal{T}^s_{i,j}u', \mathcal{M}, u' \models p$, we deduce that $\neg(u\mathcal{T}^s_{i,j}v)$. □

We characterize the accessibility relation's properties for the axioms $(4_{T,B})$ $T^s_{i,j}p \Rightarrow B_i T^s_{i,j}p$, and (S) $T^s_{i,j}p \Rightarrow B_i B_j p$.

Proposition 4.2. *For all $i, j \in \mathcal{N}$ and $(\Box, \mathcal{R}) \in \{(T^s_{i,j}, \mathcal{T}^s_{i,j}),(B_j, \mathcal{B}_j)\}$, $C \models T^s_{i,j}p \Rightarrow B_i\Box p$ if, and only if:*

$$\forall w, u, v \in \mathcal{W}, w\mathcal{B}_i u \wedge u\mathcal{R}v \Rightarrow w\mathcal{T}^s_{i,j}v$$

Proof. Let $i, j \in \mathcal{N}$ be two agents and $(\Box, \mathcal{R}) \in \{(T^s_{i,j}, \mathcal{T}^s_{i,j}),(B_j, \mathcal{B}_j)\}$,

($\Rightarrow$) By contraposition, let us suppose there exists $w, u, v \in \mathcal{W}$: $w\mathcal{B}_i u \wedge u\mathcal{R}v$ and $\neg(w\mathcal{T}^s_{i,j}v)$. Now let us define a model $\mathcal{M}$ where $i(p) = \mathcal{W} \setminus \{v\}$. Since $\neg(w\mathcal{T}^s_{i,j}v)$ then $\mathcal{M}, w \models T^s_{i,j}p$. Furthermore, as $u\mathcal{R}v$ and $\mathcal{M}, v \models \neg p$, we deduce that $\mathcal{M}, u \models \neg\Box p$ and as $w\mathcal{B}_i u$, then $\mathcal{M}, w \models \neg B_i\Box p$. We have $\mathcal{M}, w \models T^s_{i,j}p \wedge \neg B_i\Box p$. Consequently, $\not\models_C T^s_{i,j}p \Rightarrow B_i\Box p$.

($\Leftarrow$) By contraposition, $\not\models_C T^s_{i,j}p \Rightarrow B_i\Box p$. Thus, there is a model $\mathcal{M} = (\mathcal{W}, \{\mathcal{B}_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^s_{i,j}\}_{i,j\in\mathcal{N}}, i)$ and a world $w \in \mathcal{W}$ such that $\mathcal{M}, w \models T^s_{i,j}p \wedge \neg B_i\Box p$. Thus, for all $v' \in \mathcal{W}$: $w\mathcal{T}^s_{i,j}v', \mathcal{M}, v' \models p$ and there is $u \in \mathcal{W}$: $w\mathcal{B}_i u$, $\mathcal{M}, u \models \neg\Box p$. Consequently, there exists $v \in \mathcal{W}$: $u\mathcal{R}v$, $\mathcal{M}, v \models \neg p$. However, for all $v' \in \mathcal{W}$: $w\mathcal{T}^s_{i,j}v', \mathcal{M}, v' \models p$ so $\neg(w\mathcal{T}^s_{i,j}v)$. We just have proved there are $w, u, v \in \mathcal{W}$ such that $w\mathcal{B}_i u \wedge u\mathcal{R}v$ and $\neg(w\mathcal{T}^s_{i,j}v)$.

Consequently, for all $C \models T^s_{i,j}p \Rightarrow B_i\Box p$ if, and only if:

$$\forall w, u, v \in \mathcal{W}, w\mathcal{B}_i u \wedge u\mathcal{R}v \Rightarrow w\mathcal{T}^s_{i,j}v$$

□

We recall that the D axiom corresponds to the seriality property.

Proposition 4.3. *For all agents $i, j \in \mathcal{N}, C \models T^s_{i,j}p \Rightarrow \neg T^s_{i,j}\neg p$ if, and only if:*

$$\forall w \in \mathcal{W}, \exists v \in \mathcal{W} : w\mathcal{T}^s_{i,j}v$$

Proof. This is a standard proof [2]. □

Finally, we also recall the properties for all KD45 systems.

Proposition 4.4. *For all $i \in \mathcal{N}$, all KD45 axioms for $\mathcal{B}_i$ are verified in C iff C is serial, transitive and Euclidian for B_i.*

Proof. It is also a standard proof [2]. □

4.2 Soundness

Theorem 4.5. *The TB system is sound.*

Proof. (Sketch) Since we shown in the previous section that the properties of accessibility relationships in our frame preserve the validity for a formula ϕ, we just need to prove that the substitution, modus ponens and necessitation inference rules preserve the validity which are well-known theorems [2]. □

4.3 Completeness

In order to prove completeness, we define and recall firstly classical propositional theorems about maximal consistent sets, and then we define a canonical model for our axiomatic system. Finally, we prove that our canonical model satisfies each required property to preserve our axioms' validity.

4.3.1 Maximal consistent sets. In this sub-section we recall famous results about maximal consistent sets [2]:

Definition 4.6 ($\mathcal{L}_{T,B}$-inconsistency). A set Σ of formulas is *$\mathcal{L}_{T,B}$-inconsistent* iff $\exists\psi_1, ..., \psi_n \in \Sigma :\vdash \neg\bigwedge_{i=1}^n \psi_i$. A set Σ of formulas is *$\mathcal{L}_{T,B}$-consistent* iff Σ is not $\mathcal{L}_{T,B}$-inconsistent. A set of formulas Γ is *maximal $\mathcal{L}_{T,B}$-consistent* iff $\nexists\Gamma' : \Gamma \subsetneq \Gamma'$ such that Γ' is $\mathcal{L}_{T,B}$-consistent.

We recall Lindenbaum's lemma that will allow us to demonstrate our completeness theorem.

Lemma 4.7 (Lindenbaum's lemma). *For all sets Γ which are $\mathcal{L}_{T,B}$-consistent, there exists a set of formulas Γ' such that $\Gamma \subseteq \Gamma'$ and Γ' maximal $\mathcal{L}_{T,B}$-consistent.*

Finally, we recall some important properties about maximal $\mathcal{L}_{T,B}$-consistent sets.

Proposition 4.8. *For all Γ maximal $\mathcal{L}_{T,B}$-consistent and $\phi, \psi \in \mathcal{L}_{TB}$ two formulas.*

1. *MCS1:* $\Gamma \vdash \phi \Longrightarrow \phi \in \Gamma$
2. *MCS2:* $(\phi \in \Gamma \vee \neg\phi \in \Gamma) \wedge \neg(\phi \in \Gamma \wedge \neg\phi \in \Gamma)$
3. *MCS3:* $(\phi \vee \psi \in \Gamma) \Longleftrightarrow \phi \in \Gamma$ *or* $\psi \in \Gamma$
4. *MCS3':* $(\phi \wedge \psi \in \Gamma) \Longleftrightarrow \phi \in \Gamma$ *and* $\psi \in \Gamma$
5. *MCS4:* $[(\phi \Rightarrow \psi \in \Gamma) \wedge (\phi \in \Gamma)] \Longrightarrow \psi \in \Gamma$
6. *MCS5:* $\vdash \phi$ *iff* $\forall\Gamma'$ *maximal $\mathcal{L}_{T,B}$-consistent, $\phi \in \Gamma'$*

4.3.2 Canonical model. A canonical model allows to make the direct correspondence between a theorem of our system and the validity of a formula in this model. A model $\mathcal{M}^c$ is a *canonical model* of our system TB if it satisfies the following definition:

Definition 4.9 (Canonical model). Let $\mathcal{M}^c = (\mathcal{W}^c, \{\mathcal{B}^c_i\}_{i\in\mathcal{N}}, \{\mathcal{T}^c_{i,j}\}_{i,j\in\mathcal{N}}, i^c)$ be a Kripke model on $\mathcal{L}_{T,B}$ such that:

- $\mathcal{W}^c$ is a non-empty set of worlds where each world is a maximal $\mathcal{L}_{T,B}$-consistent set of formulas,

- $\{\mathcal{B}_i^c\}_{i\in\mathcal{N}}$ is a set of binary relations such that:

$$\forall i \in \mathcal{N}, \forall w \in \mathcal{W} : w\mathcal{B}_i^c v \text{ iff } B_i\phi \in w \Rightarrow \phi \in v$$

- $\{\mathcal{T}_{i,j}^c\}_{i,j\in\mathcal{N}}$ is a set of binary relations such that:

$$\forall i,j \in \mathcal{N}, \forall w \in \mathcal{W}^c : w\mathcal{T}_{i,j}^c v \text{ iff } T_{i,j}^s\phi \in w \Rightarrow \phi \in v$$

- $i^c : \mathcal{P} \to 2^{\mathcal{W}}$ is an interpretation function such that:

$$\forall p \in \mathcal{P}, w \in i^c(p) \text{ iff } p \in w$$

We consider the following notations:

- $\forall i,j \in \mathcal{N}, \forall w \in \mathcal{W}^c, \mathcal{T}_{i,j}^*(w) := \{\phi \in \mathcal{L}_{T,B} | T_{i,j}^s\phi \in w\}$
- $\forall i \in \mathcal{N}, \forall w \in \mathcal{W}^c, \mathcal{B}_i^*(w) := \{\phi \in \mathcal{L}_{T,B} | B_i\phi \in w\}$

The relations $\mathcal{T}_{i,j}^c$ and $\mathcal{B}_i^c$ become:

- $\forall i,j \in \mathcal{N}, \forall w,v \in \mathcal{W}^c : w\mathcal{T}_{i,j}^c v$ iff $\mathcal{T}_{i,j}^*(w) \subseteq v$
- $\forall i \in \mathcal{N}, \forall w,v \in \mathcal{W}^c : w\mathcal{B}_i^c v$ iff $\mathcal{B}_i^*(w) \subseteq v$

4.3.3 Canonical model and axiomatic system. Let us consider $\mathcal{M}^c = (\mathcal{W}^c, \{\mathcal{B}_i^c\}_{i\in\mathcal{N}}, \{\mathcal{T}_{i,j}^c\}_{i,j\in\mathcal{N}}, i^c)$ a canonical model of TB.

Lemma 4.10. *Let $i,j \in \mathcal{N}$ and $\phi \in \mathcal{L}_{T,B}$,*

- $\forall w \in \mathcal{W}^c : \neg T_{i,j}^s\phi \in w \Rightarrow \mathcal{T}_{i,j}^*(w) \cup \{\neg\phi\}$ *is $\mathcal{L}_{T,B}$-consistent.*
- $\forall w \in \mathcal{W}^c : \neg\mathcal{B}_i\phi \in w \Rightarrow B_i^*(w) \cup \{\neg\phi\}$ *is $\mathcal{L}_{T,B}$-consistent.*

Proof. Let $w \in \mathcal{W}^c$, $i,j \in \mathcal{N}$, $(\Box, \mathcal{R}) \in \{(T_{i,j}^s, \mathcal{T}_{i,j}^c), (B_i, \mathcal{B}_i^c)\}$. Let us assume by contraposition that $\mathcal{R}^*(w) \cup \{\neg\phi\}$ is $\mathcal{L}_{T,B}$-inconsistent. Thus, there exists $n \in \mathbb{N}$ and $\psi_1, \ldots, \psi_n \in \mathcal{R}^*(w)$ such that:

(1) $\vdash \neg(\bigwedge_{k=1}^n \psi_k \wedge \neg\phi)$
(2) $\vdash \neg \bigwedge_{k=1}^n \psi_k \vee \neg\neg\phi$
(3) $\vdash \bigwedge_{k=1}^n \psi_k \Rightarrow \phi$
(4) $\vdash \Box(\bigwedge_{k=1}^n \psi_k \Rightarrow \phi)$
(5) $\vdash (\Box \bigwedge_{k=1}^n \psi_k \Rightarrow \Box\phi)$
(6) $\vdash (\bigwedge_{k=1}^n \Box\psi_k \Rightarrow \Box\phi)$
(7) $\vdash \neg(\bigwedge_{k=1}^n \Box\psi_k \wedge \neg\Box\phi)$

Consequently, $\{\Box\psi_1, \ldots, \Box\psi_n, \neg\Box\phi\}$ is $\mathcal{L}_{T,B}$-inconsistent. However, $\forall k \in [|1, \ldots, n|]$, $\psi_k \in \mathcal{R}^*(w)$ if, and only if, $\Box\psi_k \in w$ and w is maximal $\mathcal{L}_{T,B}$-consistent. Thus, $\bigwedge_{k=1}^n \Box\psi_k \in w$ (MCS3') and then $\{\Box\psi_1, \ldots, \Box\psi_n\}$ is $\mathcal{L}_{T,B}$-inconsistent. As $\{\Box\psi_1, \ldots, \Box\psi_n\} \cup \{\neg\Box\phi\}$ is $\mathcal{L}_{T,B}$-inconsistent, $\neg\Box\phi$ does not belong to a maximal $\mathcal{L}_{T,B}$-consistent set. Consequently, $\neg\Box\phi \notin w$[1].

Thus, we proved that if $\neg\Box\phi \in w$, then $\mathcal{R}^*(w) \cup \{\neg\phi\}$ is $\mathcal{L}_{T,B}$-consistent. □

We need a third lemma to demonstrate the completeness of our system.

Lemma 4.11. *Let $w \in \mathcal{W}^c$ and $\phi \in \mathcal{L}_{T,B}$,*

$$\mathcal{M}^c, w \models \phi \textit{ iff } \phi \in w$$

Proof. Let us demonstrate the lemma by induction on the degree $n \in \mathbb{N}$ of a formula.

(Initialisation) If $\phi \in \mathcal{L}_{T,B}$ is a 0-degree formula, there exists $p \in \mathcal{P}, \phi = p$. By definition of the canonical model, we have $\forall w \in \mathcal{W}^c, w \in i(p)$ iff $p \in w$.

(Heredity) For all formulas $\phi \in \mathcal{L}_{T,B}$ of degree $< n$ with $n \in \mathbb{N}$ and for all $w \in \mathcal{W}^c$: $\mathcal{M}^c, w \models \phi$ iff $\phi \in w$.

So for all $\psi, \theta \in \mathcal{L}_{T,B}$ such that $\neg\psi$, $\psi \vee \theta$, $\psi \wedge \theta$ and $\psi \Rightarrow \theta$ are n-degree formulas for each $w \in \mathcal{W}^c$, we have (by heredity hypothesis): $\mathcal{M}^c, w \models \psi$ iff $\psi \in w$ and $\mathcal{M}^c, w \models \theta$ iff $\theta \in w$. It is standard to show heredity holds for each formula [2].

Let $(\mathcal{R}, \Box) \in \{(\mathcal{B}_i, B_i), (\mathcal{T}_{i,j}^s, T_{i,j}^s)\}$, $w \in \mathcal{W}^c$ and $\Box\psi$ a n-degree formula.

($\Rightarrow$) By contraposition let us assume that $\Box\psi \notin w$ and, as w is maximal $\mathcal{L}_{T,B}$-consistent, we have $\neg\Box\psi \in w$. By Lemma 4.10, we deduce $\mathcal{R}^*(w) \cup \{\neg\psi\}$ is $\mathcal{L}_{T,B}$-consistent. By Lemma 4.7, we deduce there exists a $v \in \mathcal{W}^c : \mathcal{R}^*(w) \cup \{\neg\psi\} \subseteq v$ and v is maximal $\mathcal{L}_{T,B}$-consistent.

Thus, $\neg\psi \in v$ and, by definition of $\mathcal{R}^c$, we have $w\mathcal{R}^c v$ and $\psi \notin v$. By the induction hypothesis, we have $\mathcal{M}^c, v \not\models \psi$. Since there exists $v \in \mathcal{W}^c : w\mathcal{R}^c v : v \models \neg\psi$, we have $\mathcal{M}^c, w \models \neg\Box\psi$, i.e, $\mathcal{M}^c, w \not\models \Box\psi$.

($\Leftarrow$) By contraposition, let us assume that $\mathcal{M}^c, w \not\models \Box\psi$, i.e. $\mathcal{M}^c, w \models \neg\Box\psi$. Thus, there exists $v \in \mathcal{W} : w\mathcal{R}^c v, \mathcal{M}^c, v \models \neg\psi$. Consequently, $\mathcal{M}^c, v \not\models \phi$ and, by the induction hypothesis, we have $\phi \notin v$. However, since $\phi \notin v$, by definition of $\mathcal{R}^c$, we deduce that $\Box\phi \notin w$.

(Conclusion)

$$\forall \phi \in \mathcal{L}_{T,B}, \forall w \in \mathcal{W}^c : \mathcal{M}^c, w \models \phi \text{ iff } \phi \in w$$

□

Now, let us prove the connection between our canonical model and the formulas proved by our system.

Proposition 4.12. *Let $\phi \in \mathcal{L}_{T,B}$,*

$$\mathcal{M}^c \models \phi \textit{ iff } \vdash \phi$$

Proof. (1) By definition $\mathcal{M}^c \models \phi$ iff $\forall w \in \mathcal{W}^c : \mathcal{M}^c, w \models \phi$
(2) By Lemma 4.11 $\forall w \in \mathcal{W}^c : \mathcal{M}^c, w \models \phi$ iff $\forall w \in \mathcal{W}^c, \phi \in w$
(3) Finally by MCS5, $\forall w \in \mathcal{W}^c, \phi \in w$ iff $\vdash \phi$.
Consequently, $\mathcal{M}^c \models \phi$ iff $\vdash \phi$. □

4.3.4 Completeness proof. Now that we have recalled main results about canonical models, we are able to prove the completeness.

Lemma 4.13. *Let $\mathcal{M}^c = (\mathcal{W}^c, \{\mathcal{B}_i^c\}_{i\in\mathcal{N}}, \{\mathcal{T}_{i,j}^c\}_{i,j\in\mathcal{N}}, i^c)$ be a canonical model for TB. We have:*

(1) $\forall i,j \in \mathcal{N}, \mathcal{T}_{i,j}^c$ *is serial*
(2) $\forall i,j \in \mathcal{N}, \forall w,u,v \in \mathcal{W}^c, w\mathcal{B}_i^c u \wedge w\mathcal{T}_{i,j}^c v \Rightarrow u\mathcal{T}_{i,j}^c v$
(3) $\forall i,j \in \mathcal{N}, \forall w,u,v \in \mathcal{W}^c, w\mathcal{B}_i^c u \wedge u\mathcal{T}_{i,j}^c v \Rightarrow w\mathcal{T}_{i,j}^c v$
(4) $\forall i,j \in \mathcal{N}, \forall w,u,v \in \mathcal{W}^c, w\mathcal{B}_i^c u \wedge u\mathcal{B}_j^c v \Rightarrow w\mathcal{T}_{i,j}^c v$
(5) $\forall i \in \mathcal{N}, \mathcal{B}_i^c$ *is serial, transitive and Euclidean*

Proof. Let $i,j \in \mathcal{N}$, $w \in \mathcal{W}^c$ and $T_{i,j}^s\phi \in w$.

(1) This is a standard proof of KD completeness [2].

(2) Let $i,j \in \mathcal{N}$. For all $w,u,v \in \mathcal{W}^c : w\mathcal{B}_i^c u \wedge w\mathcal{T}_{i,j}^c v$ and $\phi \notin v$. By MCS2, $\neg\phi \in v$ and, since $w\mathcal{T}_{i,j}^c v$, we have $\neg T_{i,j}^s\phi \in w$. However $\vdash \neg T_{i,j}^s\phi \Rightarrow B_i\neg T_{i,j}^s\phi$. Thus by MCS5, we have $\neg T_{i,j}^s\phi \Rightarrow B_i\neg T_{i,j}^s\phi \in w$. Moreover by MCS4, we deduce that $B_i\neg T_{i,j}^s\phi \in w$ and since $w\mathcal{B}_i^c u$, we have $\neg T_{i,j}^s\phi \in u$. Then by MCS2, $T_{i,j}^s\phi \notin u$. By contraposition, we have $T_{i,j}^s\phi \in u \Rightarrow \phi \in v$, and thus $u\mathcal{T}_{i,j}^c v$.

(3) Let $i,j \in \mathcal{N}$. For all $w,u,v \in \mathcal{W}^c : w\mathcal{B}_i^c u \wedge u\mathcal{T}_{i,j}^c v$ and $T_{i,j}^s\phi \in w$. However, $\vdash T_{i,j}^s\phi \Rightarrow B_iT_{i,j}^s\phi$. Thus by MCS5, $T_{i,j}^s\phi \Rightarrow$

[1] If we had $\neg\Box\phi \in w$, we would also have $\bigwedge_{k=1}^n \Box\psi_k \wedge \neg\Box\phi \in w$ (by MCS3') and then $\{\Box\psi_1, \ldots, \Box\psi_n, \neg\Box\phi\}$ would be $\mathcal{L}_{T,B}$-consistent, which is a contradiction.

$B_i T^s_{i,j}\phi \in w$ and, by MCS4, $B_i T^s_{i,j}\phi \in w$. Consequently $T^s_{i,j}\phi \in u$, and then $\phi \in v$. Thus, by definition of $\mathcal{T}^c_{i,j}$, we have $w\mathcal{T}^c_{i,j}v$.

(4) Let $i, j \in \mathcal{N}$. For all $w, u, v \in \mathcal{W}^c : w\mathcal{B}^c_i u \wedge u\mathcal{B}^c_j v$ and $T^s_{i,j}\phi \in w$. However, $\vdash T^s_{i,j}\phi \Rightarrow B_i B_j \phi$. Thus by MCS5 $T^s_{i,j}\phi \Rightarrow B_i B_j \phi \in w$, and, by MCS4, $B_i B_j \phi \in w$. Consequently $B_j \phi \in u$, and then $\phi \in v$. Thus, by definition of $\mathcal{T}^c_{i,j}$, we have $w\mathcal{T}^c_{i,j}v$, i.e. we shown that:

$$\forall i, j \in \mathcal{N}, \forall w, u, v \in \mathcal{W}^c, w\mathcal{B}^c_i u \wedge u\mathcal{B}^c_j v \Rightarrow w\mathcal{T}^c_{i,j}v$$

(5) This is a standard proof of KD45 completeness [2]. □

Theorem 4.14. *The TB system is complete.*

Proof. By synthesis, we have:

(1) $\mathcal{C} \models \phi \Longrightarrow \mathcal{M}^c \models \phi$
(2) $\mathcal{M}^c \models \phi \Longleftrightarrow \vdash \phi$

Consequently, $\mathcal{C} \models \phi \Longrightarrow \vdash \phi$. □

5 PROPERTIES

In this section, we show some interesting properties.

5.1 Trust in the sincerity is distributive

As we consider a normal logic, we have the following properties:

Proposition 5.1. *Let $i, j \in \mathcal{N}$.*

(1) $\vdash T^s_{i,j}\phi \wedge T^s_{i,j}\psi \equiv T^s_{i,j}(\phi \wedge \psi) \quad (\wedge_T)$
(2) $\vdash (T^s_{i,j}\phi \vee T^s_{i,j}\psi) \Rightarrow T^s_{i,j}(\phi \vee \psi) \quad (\vee_T)$

Proof. Since $T^s_{i,j}$ is a normal modality we immediately deduce these properties [2]. □

Indeed, an agent i cannot trust an inconsistent discourse (the set of propositions formulated by the agent j) because in our TB system this would lead it to trust in any proposal from j.

5.2 Some belief-related properties

The reciprocals of axioms $(4_{T,B})$ and $(5_{T,B})$ hold.

Proposition 5.2. *Let $i, j \in \mathcal{N}$ be two agents,*

(1) $\vdash B_i T^s_{i,j}p \Rightarrow T^s_{i,j}p \quad (C4_{T,B})$
(2) $\vdash B_i \neg T^s_{i,j}p \Rightarrow \neg T^s_{i,j}p \quad (C5_{T,B})$

Proof. Let $i, j \in \mathcal{N}$ be two agents. We prove the first property:

(1) $\vdash \neg T^s_{i,j}p \Rightarrow B_i \neg T^s_{i,j}p \quad (5_{T,B})$
(2) $\vdash B_i \neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p \quad (D_B)$
(3) $\vdash (\neg T^s_{i,j}p \Rightarrow (B_i \neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p))$
(4) $\vdash (\neg T^s_{i,j}p \Rightarrow (B_i \neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p)) \Rightarrow$
$((\neg T^s_{i,j}p \Rightarrow B_i \neg T^s_{i,j}p) \Rightarrow (\neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p))$
(5) $\vdash \neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p$
(6) $\vdash (\neg T^s_{i,j}p \Rightarrow \neg B_i T^s_{i,j}p) \Rightarrow (B_i T^s_{i,j}p \Rightarrow T^s_{i,j}p)$
(7) $\vdash B_i T^s_{i,j}p \Rightarrow T^s_{i,j}p$

We prove the second property:

(1) $\vdash B_i \neg T^s_{i,j}p \Rightarrow \neg B_i T_{i,j}p \quad (D_B)$
(2) $\vdash T^s_{i,j}p \Rightarrow B_i T^s_{i,j}p \quad (4_{T,B})$
(3) $\vdash (T^s_{i,j}p \Rightarrow B_i T^s_{i,j}p) \Rightarrow (\neg B_i T^s_{i,j}p \Rightarrow \neg T^s_{i,j}p)$
(4) $\vdash \neg B_i T^s_{i,j}p \Rightarrow \neg T^s_{i,j}p$
(5) $\vdash (B_i \neg T^s_{i,j}p \Rightarrow (\neg B_i T^s_{i,j}p \Rightarrow \neg T_{i,j}p)$
(6) $\vdash (B_i \neg T^s_{i,j}p \Rightarrow (\neg B_i T^s_{i,j}p \Rightarrow \neg T_{i,j}p) \Rightarrow$
$((B_i \neg T^s_{i,j}p) \Rightarrow (\neg B_i T^s_{i,j}p)) \Rightarrow ((B_i \neg T^s_{i,j}p) \Rightarrow (\neg T_{i,j}p))$
(7) $\vdash B_i \neg T_{i,j}p \Rightarrow \neg T_{i,j}p$

□

Those properties highlights (1) when it is the case agents believe they trust, then it is the case they trust; (2) when it is the case they believe they do not trust, then it is the case they do not. Finally we consider a last belief-related property.

Proposition 5.3. *For all agents $i, j \in \mathcal{N}$,*

$$\vdash B_i B_j \phi \Rightarrow \neg T^s_{i,j} \neg \phi$$

Proof. Let $i, j \in \mathcal{N}$.

(1) $\vdash B_j \phi \Rightarrow \neg B_j \neg \phi$
(2) $\vdash B_i(B_j \phi \Rightarrow \neg B_j \neg \phi)$
(3) $\vdash B_i(B_j \phi \Rightarrow \neg B_j \neg \phi) \Rightarrow (B_i B_j \phi \Rightarrow B_i \neg B_j \neg \phi)$
(4) $\vdash B_i B_j \phi \Rightarrow B_i \neg B_j \neg \phi$
(5) $\vdash B_i \neg B_j \neg \phi \Rightarrow \neg B_i B_j \neg \phi) \Rightarrow \neg B_i B_j \neg \phi$
(6) $\vdash T^s_{i,j} \neg \phi \Rightarrow B_i B_j \neg \phi$
(7) $\vdash (T^s_{i,j} \neg \phi \Rightarrow B_i B_j \neg \phi) \Rightarrow (\neg B_i B_j \neg \phi \Rightarrow \neg T^s_{i,j} \neg \phi)$
(8) $\vdash \neg B_i B_j \neg \phi \Rightarrow \neg T^s_{i,j} \neg \phi$
(9) $\vdash (B_i B_j \phi \Rightarrow B_i \neg B_j \neg \phi \Rightarrow \neg B_i B_j \neg \phi) \Rightarrow$
$((B_i B_j \phi \Rightarrow B_i \neg B_j \neg \phi) \Rightarrow (B_i B_j \phi \Rightarrow \neg B_i B_j \neg \phi))$
(10) $\vdash (B_i B_j \phi \Rightarrow \neg B_i B_j \neg \phi \Rightarrow \neg T^s_{i,j} \neg \phi) \Rightarrow$
$(((B_i B_j \phi \Rightarrow \neg B_i B_j \neg \phi) \Rightarrow (B_i B_j \phi \Rightarrow \neg T^s_{i,j} \neg \phi))$
(11) $\vdash B_i B_j \phi \Rightarrow \neg T^s_{i,j} \neg \phi$

□

5.3 Trust in the sincerity is not transitive

Some studies have already pointed out reasons why trust was not transitive [3]. Trust in the sincerity is not transitive either. By transitivity, we mean that we do not have an inference rule deducing that if $T^s_{i,j} T^s_{j,k} \phi$ then $T^s_{i,k} \phi$. Indeed, it is not because an agent i trusts in the sincerity of an agent j when j states that it trusts in the sincerity of another agent k that the agent i necessarily trusts in the sincerity of k for this same proposition, as j may be sincere and nevertheless be wrong. However, the following property may be interesting as pseudo-transitivity:

Proposition 5.4. *For all agents $i, j, k \in \mathcal{N}$,*

$$\vdash T^s_{i,j} T^s_{j,k} \phi \Rightarrow B_i B_j B_k \phi$$

Proof. Let $i, j, k \in \mathcal{N}$.

(1) $\vdash T^s_{i,j} T^s_{j,k} \phi \Rightarrow B_i B_j T^s_{j,k} \phi$
(2) $\vdash T^s_{j,k} \phi \Rightarrow B_j B_k \phi$
(3) $\vdash B_j T^s_{j,k} \phi \Rightarrow T^s_{j,k} \phi$
(4) $\vdash B_i(B_j T^s_{j,k} \phi \Rightarrow T^s_{j,k} \phi)$
(5) $\vdash B_i(B_j T^s_{j,k} \phi \Rightarrow T^s_{j,k} \phi) \Rightarrow B_i B_j T^s_{j,k} \phi \Rightarrow B_i T^s_{j,k} \phi$
(6) $\vdash B_i B_j T^s_{j,k} \phi \Rightarrow B_i T^s_{j,k} \phi$
(7) $\vdash T^s_{i,j} T^s_{j,k} \phi \Rightarrow B_i B_j B_k \phi$

□

5.4 Shared trust

We can extend our notion of trust to groups of agents in order to express *shared trust*. Let us remark that we restrict ourselves to this notion as a first approach. Other aspects of collective trust, such as *reciprocal trust* or *mutual trust*, are interesting but are left for future works.

5.4.1 Definition. To define *shared trust*, we rely on the definition of Smith *et al.* [23]: a group of agents trusts another group of agents if, and only if, all agents of the first group trust all agents of the second group.

$$\forall I, J \subseteq \mathcal{N} : Tc_{I,J}\phi \triangleq \bigwedge_{(i,j)\in I\times J} T^s_{i,j}\phi$$

This is a consensus in the sense that all agents of I must trust all agents of J with respect to the same statement. Moreover, we consider a dual notion of shared trust, denoted by $Tc^*_{I,J}$, as follows:

$$\forall I, J \subseteq \mathcal{N} : Tc^*_{I,J}\phi \triangleq \bigvee_{(i,j)\in I\times J} T^s_{i,j}\phi$$

This predicate expresses that at least one agent of I trusts another agent of J. Indeed, if no agent of I trusts the agents of J for ϕ then $\neg Tc^*_{I,J}\phi$. Let us remark that shared trust may be defined differently in the literature. For instance, Herzig *et al.* [10] consider a *reputation* predicate indicating that a *majority* of agents of I has a dispositional trust towards the agents of J. For the sake of simplicity, we do not introduce a notion of a majority and therefore we do not consider this notion of reputation.

5.4.2 Shared trust behaves like a KD system. Shared trust has the following properties:

Proposition 5.5. *For all $I, J, K \subseteq \mathcal{N}$:*

(1) $\vdash Tc_{I,J}\phi \wedge Tc_{I,J}\psi \equiv Tc_{I,J}(\phi \wedge \psi)$
(2) $\vdash (Tc_{I,J}\phi \vee Tc_{I,J}\psi) \Rightarrow Tc_{I,J}(\phi \vee \psi)$
(3) $\vdash (Tc_{I,J}\phi \wedge Tc_{I,J}(\phi \Rightarrow \psi)) \Rightarrow Tc_{I,J}\psi$
(4) $\vdash Tc_{I,J}\phi \Rightarrow \neg Tc^*_{I,J}\neg\phi$
(5) $\vdash Tc_{I,J}\phi \Rightarrow \neg Tc_{I,J}\neg\phi$

Proof. (Sketches) For all $I, J, K \subseteq \mathcal{N}$,

(1) $\vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \wedge T^s_{i,j}\psi) \equiv \bigwedge_{(i,j)\in I\times J} T^s_{i,j}(\phi \wedge \psi)$
(2) $\vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \vee T^s_{i,j}\psi) \Rightarrow \bigwedge_{(i,j)\in I\times J} T^s_{i,j}(\phi \vee \psi)$
(3) is obtained by :

- $\{Tc_{I,J}\phi \wedge Tc_{I,J}(\phi \Rightarrow \psi)\} \vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \wedge (T^s_{i,j}(\phi \Rightarrow \psi)))$
- $\{Tc_{I,J}\phi \wedge Tc_{I,J}(\phi \Rightarrow \psi)\} \vdash \bigwedge_{(i,j)\in I\times J} T^s_{i,j}\psi$

Consequently, $\vdash (Tc_{I,J}\phi \wedge Tc_{I,J}(\phi \Rightarrow \psi)) \Rightarrow Tc_{I,J}\psi$.

(4) is obtained by :

- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \wedge (T^s_{i,j}\phi \Rightarrow \neg T^s_{i,j}\neg\phi))$
- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J} \neg T^s_{i,j}\neg\phi$
- $\{Tc_{I,J}\phi\} \vdash \neg \bigvee_{(i,j)\in I\times J} T^s_{i,j}\neg\phi$

Consequently, $\vdash Tc_{I,J}\phi \Rightarrow \neg Tc^*_{I,J}\neg\phi$.

(5) is obtained by :

- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \wedge (T^s_{i,j}\phi \Rightarrow \neg T^s_{i,j}\neg\phi))$
- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J} \neg T^s_{i,j}\neg\phi$
- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J} \neg T^s_{i,j}\neg\phi \Rightarrow \bigvee_{(i,j)\in I\times J} \neg T^s_{i,j}\neg\phi$
- $\{Tc_{I,J}\phi\} \vdash \bigvee_{(i,j)\in I\times J} \neg T^s_{i,j}\neg\phi \Rightarrow \neg \bigwedge_{(i,j)\in I\times J} T^s_{i,j}\neg\phi$
- $\{Tc_{I,J}\phi\} \vdash \neg \bigwedge_{(i,j)\in I\times J} T^s_{i,j}\neg\phi$

Consequently, $\vdash Tc_{I,J}\phi \Rightarrow \neg Tc_{I,J}\neg\phi$. □

Hence, shared trust behaves like a KD system: the trust in the sincerity axiomatics is the same as the shared trust level.

5.4.3 Shared trust implies common beliefs.

Proposition 5.6. *For all $I, J, K \subseteq \mathcal{N}$,*

$$(1) \vdash Tc_{I,J}\phi \Rightarrow \bigwedge_{(i,j)\in I\times J} B_iB_j\phi$$

$$(2) \vdash Tc_{I,J}Tc_{J,K}\phi \Rightarrow \bigwedge_{(i,j,k)\in I\times J\times K} B_iB_jB_k\phi$$

Proof. (Sketches) For all $I, J, K \subseteq \mathcal{N}$

(1) is obtained by :

- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j}\phi \wedge (T^s_{i,j}\phi \Rightarrow B_iB_j\phi))$
- $\{Tc_{I,J}\phi\} \vdash \bigwedge_{(i,j)\in I\times J} B_iB_j\phi$

Consequently, $\vdash Tc_{I,J}\phi \Rightarrow \bigwedge_{(i,j)\in I\times J} B_iB_j\phi$.

(2) is obtained by :

- $\{Tc_{I,J}Tc_{J,K}\phi\} \vdash \bigwedge_{(i,j)\in I\times J}(T^s_{i,j} \bigwedge_{k\in K} T_{j,k}\phi \wedge (T^s_{i,j}T^s_{j,k}\phi \Rightarrow B_iB_jB_k\phi))$
- $\{Tc_{I,J}Tc_{J,K}\phi\} \vdash \bigwedge_{(i,j,k)\in I\times J\times K} B_iB_jB_k\phi$

Consequently, $\vdash Tc_{I,J}Tc_{J,K}\phi \Rightarrow \bigwedge_{(i,j,k)\in I\times J\times K} B_iB_jB_k\phi$. □

Let us notice that those proofs work because of $(\wedge_T)$ and $\forall k \in \mathcal{N}$, $\vdash B_k(p \wedge q) \equiv B_kp \wedge B_kq$. Thanks to those properties, we show that if two groups trust in the sincerity of the other, it implies that each agent of I believes that each other agent of J believes what it says.

6 CONCLUSION AND PERSPECTIVES

To conclude this article, we have proposed a normal modal logic allowing to reason about the trust in the sincerity of an agent towards another one. Considering a doxastic system, we have introduced a normal modality $T^s_{i,j}p$ meaning that an agent i trusts in the sincerity of an agent j for a proposition p. This modality allows us to consider the fact that an agent can tolerate that another is wrong since the latter did not attempt to deceive the former about p. Indeed, a direct application of this modality is to reason about trust when some agent attempt to manipulate other agents. We showed our system is sound and complete, and we exhibited some notable properties: non-transitivity of trust, shared trust as a KD system for instance. As future works, we intend to study the formal links that may exist between the reliability modality introduced by Liau [15] and ours. Furthermore, we noticed a strong connection between honesty and norm compliance as shown by Demolombe [5]. Consequently, we would like to combine our formalism with other modalities such as a deontic modality for representing norms or an action modality like those introduced by Lorini in a context of social influence [17].

REFERENCES

[1] Christiano Castelfranchi and Rino Falcone. 2010. *Trust theory: A socio-cognitive and computational model.* John Wiley & Sons.

[2] Brian F. Chellas. 1980. *Modal logic: an introduction.* Vol. 316. Cambridge University Press.

[3] Bruce Christianson and William Harbison. 1997. Why isn't trust transitive?. In *Security protocols*. 171–176.

[4] Mehdi Dastani, Andreas Herzig, Joris Hulstijn, and Leendert Van Der Torre. 2004. Inferring trust. In *5th CLIMA*. Springer, 144–160.

[5] Robert Demolombe. 2004. Reasoning about trust: A formal logical framework. In *2nd iTrust*. 291–303.

[6] Robert Demolombe and Churn-Jung Liau. 2001. A logic of graded trust and belief fusion. In *4th Workshop on Deception, Fraud and Trust in Agent Societies*. 13–25.

[7] Besik Dundua and Levan Uridia. 2010. Trust and Belief, Interrelation. In *3rd WAT*.

[8] Rino Falcone, Giovanni Pezzulo, and Cristiano Castelfranchi. 2002. A fuzzy approach to a belief-based trust computation. In *5th Workshop on Deception, Fraud and Trust in Agent Societies*. 73–86.

[9] Tuan-Fang Fan and Churn-Jung Liau. 2016. Reasoning About Justified Belief Based on the Fusion of Evidence. In *15th JELIA*. Springer, 240–255.

[10] Andreas Herzig, Emiliano Lorini, Jomi Fred Hübner, and Laurent Vercouter. 2010. A logic of trust and reputation. *Logic Journal of the IGPL* 18, 1 (2010), 214–244.

[11] Audun Josang and Roslan Ismail. 2002. The beta reputation system. In *15th Bled Electronic Commerce Conference*. 2502–2511.

[12] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. 2003. The eigentrust algorithm for reputation management in p2p networks. In *12th WWW*. ACM, 640–651.

[13] Vibhor Kant and Kamal K Bharadwaj. 2013. Fuzzy computational models of trust and distrust for enhanced recommendations. *Int. J. of Intelligent Systems* 28, 4 (2013), 332–365.

[14] Eleni Koutrouli and Aphrodite Tsalgatidou. 2011. Credibility enhanced reputation mechanism for distributed e-communities. In *19th PDP*. 627–634.

[15] Churn-Jung Liau. 2003. Belief, information acquisition, and trust in multi-agent systems - a modal logic formulation. *Artificial Intelligence* 149, 1 (2003), 31–60.

[16] Emiliano Lorini, Guifei Jiang, and Laurent Perrussel. 2014. Trust-based belief change. In *21st ECAI*. IOS Press, 549–554.

[17] Emiliano Lorini and Giovanni Sartor. 2016. A STIT Logic for Reasoning About Social Influence. *Studia Logica* 104, 4 (2016), 773–812.

[18] Guillaume Muller and Laurent Vercouter. 2005. Decentralized monitoring of agent communications with a reputation model. In *Trusting Agents for Trusting Electronic Societies*. 144–161.

[19] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The PageRank citation ranking: bringing order to the web. (1999).

[20] Yefeng Ruan and Arjan Durresi. 2016. A survey of trust management systems for online social communities – Trust modeling, trust inference and attacks. *Knowledge-Based Systems* 106 (2016), 150–163.

[21] Jordi Sabater and Carles Sierra. 2001. Regret : A reputation model for gregarious societies. In *4th Workshop on Deception, Fraud and Trust in Agent Societies*.

[22] Munindar P. Singh. 2011. Trust as dependence: A logical approach. In *10th AAMAS*. 863–870.

[23] Clara Smith, Agustín Ambrossio, Leandro Mendoza, and Antonino Rotolo. 2011. Combinations of normal and non-normal modal logics for modeling collective trust in normative MAS. In *4th AICOL*. 189–203.

[24] Thibault Vallée and Grégory Bonnet. 2015. Using KL divergence for credibility assessment. In *14th AAMAS*. 1797–1798.

[25] Jin-Long Wang and Shih-Ping Huang. 2007. Fuzzy logic based reputation system for mobile ad hoc networks. In *11th KES*. 1315–1322.

[26] Huanyu Zhao and Xiaolin Li. 2009. H-trust : A group trust management system for peer-to-peer desktop grid. *Artificial Intelligence* 24, 5 (2009), 833–843.