Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Artificial Intelligence 149 (2003) 31–60

Artificial Intelligence

www.elsevier.com/locate/artint

Belief, information acquisition, and trust in multi-agent systems—A modal logic formulation ☆

Churn-Jung Liau

Institute of Information Science, Academia Sinica, Taipei, Taiwan

Received 25 October 2001

Abstract

In this paper, we consider the influence of trust on the assimilation of acquired information into an agent's belief. By use of modal logic, we semantically and axiomatically characterize the relationship among belief, information acquisition and trust. The belief and information acquisition operators are respectively represented by KD45 and KD normal modalities, whereas trust is denoted by a modal operator with minimal semantics. One characteristic axiom of the basic system is if agent i believes that agent j has told him the truth of p and he trusts the judgement of j on p, then he will also believe p. In addition to the basic system, some variants and further axioms for trust and information acquisition are also presented to show the expressive richness of the logic. The applications of the logic to computer security and database reasoning are also suggested by its connection with some previous works.

Keywords: Trust; Belief; Information acquisition; Modal logic; Multi-agent systems; Intelligent agents

1. Introduction

In the internet age, infoglut has become a serious problem in information retrieval and search. If a keyword is input into a commonly-used search engine, it is not unusual to get back a list of thousands of web pages. Thus, the real difficulty is not how to find information, but how to find *useful* information. Recently, a great number of software agents has been designed to circumvent the problem. Agents can search the web to find and filter out information matching the user's need. However, not all internet information sources are trustworthy. Some web sites are out-of-date, some news media provide

☆ This is an expanded and revised version of [34].

E-mail address: liaucj@iis.sinica.edu.tw (C.-J. Liau).

0004-3702/$ – see front matter
doi:10.1016/S0004-3702(03)00063-8

erroneous information, and some people even intentionally spread rumors or deceive by anonymity. From the viewpoint of agent societies, each agent plays both the role of information provider and of receiver. Consequently, the information search process can be seen as the communication between two agents. A receiver has to decide whether he/she can believe the received information according to his/her trust in the provider.

In [39], an agent is characterized by mental attitudes, such as knowledge, belief, obligation, and commitment. This view of agents, in accordance with the intentional stance proposed in [19], has been widely accepted as a convenient way to analyze and describe complex systems [49]. The model of these attitudes has been the traditional concern of philosophical logic, such as epistemic logic, doxastic logic [24], and deontic logic [2]. Some logics derived from philosophical analysis have been applied to the modeling of AI and distributed systems [20,35]. In most of these logics, the mental attitudes are represented by modal operators, and their meanings are in general given by the possible world semantics for modal logic [10]. Following the approach, we would like to propose a doxastic logic with modalities for representing the trusting attitudes and the information transmission action between agents, and then discuss how one agent's belief is influenced by his/her acquired information and trust toward other agents. More specifically, in traditional doxastic logic, $B_i\varphi$ means that agent i believes φ, so we will add to the logic additional modal operators T_{ij} and I_{ij}. The intended meaning of $T_{ij}\varphi$ is that agent i trusts the judgement of j on the truth of φ, whereas $I_{ij}\varphi$ means agent i acquires information φ from j.

In the remainder of the paper, we will first give a general logic that meets the above-mentioned requirement. The syntax, semantics, and a basic axiomatic system of the logic is presented in Section 2. In Sections 3 and 4, some additional assumptions will be considered to produce variants of the basic logic. In Section 5, we compare our logic with some related works in multi-agent systems as well as in computer security and database management contexts, demonstrating the potential application of our framework beyond multi-agent systems. Finally, we conclude the paper with some perspectives for further research.

2. The basic logic BIT

The basic logic of belief, information acquisition, and trust (BIT) is an extension of the traditional doxastic logic, which is in turn a multi-agent version of the KD45 system of the normal modal logic [10]. Assume we have n agents and a set Φ_0 of countably many atomic propositions, then the set of well-formed formulas (wff) for the logic BIT is the least set containing Φ_0 and closed under the following formation rules:

- if φ is a wff, so are $\neg\varphi$, $B_i\varphi$, $I_{ij}\varphi$, and $T_{ij}\varphi$ for all $1 \leqslant i \neq j \leqslant n$, and
- if φ and ψ are wffs, then $\varphi \vee \psi$ is, too.

As usual, other classical Boolean connectives $\wedge$ (and), $\supset$ (implication), $\equiv$ (equivalence), $\top$ (tautology), and $\perp$ (contradiction) can be defined as abbreviations.

The possible-worlds semantics provides a general framework for the modeling of knowledge and belief [20]. In the semantics, an agent's belief state corresponds to the

extent to which he can determine what world he is in. In a given world, the belief state determines the set of worlds that the agent considers possible. Then an agent is said to believe a fact φ if φ is true in all worlds in this set. Analogously, the information of an agent acquired from another agent constrains the possibility of the worlds according to the acquired information. However, since an agent perceives the possibility that other agents may be unreliable, he will not blindly believe all acquired information. Thus, the set of possible worlds according to acquired information from some particular agent may be different from that associated with his belief state. Of course, since an agent may lie, the information of other agents acquired from him may not be compatible with what he believes. On the other hand, the semantics of trust is relatively more "syntactic" and less restrictive. Though trust in general depends on some rational factors such as the honesty and credibility of the trusted agent, it also usually contains some irrational or emotional components. Since the assessment of credibility of an agent can only depend on his past records, we can not guarantee that the agent does not provide any wrong information in the future. Even very respectable news media may make some errors, so any trust must be accompanied with risk. This means that we will only impose minimal constraint on the set of statements on which an agent trusts another agent's judgement.

From the semantics, the trust operators do not have logical closure property whereas the belief and information acquisition operators do. The asymmetry arises from our modeling of agents as perfect reasoners. The beliefs of a perfect reasoner are closed under logical consequence and, since an agent can reason from the acquired information, when he receives φ from another agent, he also implicitly receives the consequences implied by φ. However, trust is a totally different thing. When an agent trusts another on φ, he does not necessarily trust the latter on all consequences derived from φ, even if he is aware of all these consequences due to his reasoning capability. A real example to illustrate the phenomenon is given in Section 3 (Example 1).

According to the informal discussion above, the formal semantics for B_i and I_{ij} is the Kripke semantics for normal modal operators, whereas that for T_{ij} is the so-called minimal (or neighborhood) semantics [10]. Formally, a BIT model is a tuple $(W, \pi, (\mathcal{B}_i)_{1\leqslant i\leqslant n}, (\mathcal{I}_{ij})_{1\leqslant i\neq j\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n})$, where

- W is a set of possible worlds,
- $\pi : \Phi_0 \to 2^W$ is a truth assignment mapping each atomic proposition to the set of worlds in which it is true,
- $\mathcal{B}_i \subseteq W \times W$ is a serial, transitive and Euclidean binary relation[1] on W,
- $\mathcal{I}_{ij} \subseteq W \times W$ is a serial relation on W,
- $\mathcal{T}_{ij} \subseteq W \times 2^W$ is a binary relation between W and the power set of W.

In the following, we will use some standard notations for binary relations. If $\mathcal{R} \subseteq A \times B$ is a binary relation between A and B, we will write $\mathcal{R}(a, b)$ for $(a, b) \in \mathcal{R}$ and $\mathcal{R}(a)$ for the subset $\{b \in B \mid \mathcal{R}(a, b)\}$. Thus for any $w \in W$, $\mathcal{B}_i(w)$ and $\mathcal{I}_{ij}(w)$ will be subsets of

[1] A relation $\mathcal{R}$ on W is serial if $\forall w \exists u \mathcal{R}(w, u)$, transitive if $\forall w, u, v(\mathcal{R}(w, u) \wedge \mathcal{R}(u, v) \Rightarrow \mathcal{R}(w, v))$, and Euclidean if $\forall w, u, v(\mathcal{R}(w, u) \wedge \mathcal{R}(w, v) \Rightarrow \mathcal{R}(u, v))$.

W, whereas $\mathcal{T}_{ij}(w)$ is a subset of 2^W. Informally, $\mathcal{B}_i(w)$ is the set of worlds that agent i considers possible under w according to his belief, whereas $\mathcal{I}_{ij}(w)$ is what the agent i considers possible according to the information acquired from j. On the other hand, since each subset of W is the semantic counterpart of a proposition, for any $S \subseteq W$, $S \in \mathcal{T}_{ij}(w)$ means that agent i trust j's judgement on the truth of the proposition corresponding to S. The informal intuition is reflected in our formal definition of satisfaction relation. Let $M = (W, \pi, (\mathcal{B}_i)_{1\leqslant i\leqslant n}, (\mathcal{I}_{ij})_{1\leqslant i\neq j\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n})$ be a BIT model and Φ be the set of wffs, then the satisfaction relation $\models_M \subseteq W \times \Phi$ is defined by the following inductive rules (we will use the infix notation for the relation and omit the subscript M for convenience):

(1) $w \models p$ iff $w \in \pi(p)$ when $p \in \Phi_0$,
(2) $w \models \neg\varphi$ iff $w \not\models \varphi$,
(3) $w \models \varphi \vee \psi$ iff $w \models \varphi$ or $w \models \psi$,
(4) $w \models B_i\varphi$ iff for all $u \in \mathcal{B}_i(w)$, $u \models \varphi$,
(5) $w \models I_{ij}\varphi$ iff for all $u \in \mathcal{I}_{ij}(w)$, $u \models \varphi$,
(6) $w \models T_{ij}\varphi$ iff $|\varphi| \in \mathcal{T}_{ij}(w)$, where $|\varphi| = \{u \in W\colon u \models \varphi\}$ is called the truth set of φ.

As usual, we can define validity from the satisfaction relation. A wff φ is valid in M, denoted by $\models_M \varphi$, if $|\varphi| = W$. Let $\mathbf{C}$ be a class of BIT models, then $\models_{\mathbf{C}} \varphi$ if for all $M \in \mathbf{C}$, we have $\models_M \varphi$. Let $\Sigma \cup \{\varphi\} \subseteq \Phi$, then $\Sigma \models_{\mathbf{C}} \varphi$ denotes that for all $M \in \mathbf{C}$ and w in M, if for all $\psi \in \Sigma$, $w \models_M \psi$ then $w \models_M \varphi$.

So far, we have defined a BIT model so that the relations $\mathcal{B}_i$, $\mathcal{I}_{ij}$, and $\mathcal{T}_{ij}$ are completely independent. This means that the information an agent acquired from other agents may be completely irrelevant to his belief, so the agent will not benefit from communication with others. This is definitely not what we want to model. Though we do not want an agent to believe blindly what other agents tell him, it is indeed inevitable that his belief should be influenced by the information he acquired from agents he trusts. Based on this consideration, we will impose some constraints on the BIT models. Let $M = (W, \pi, (\mathcal{B}_i)_{1\leqslant i\leqslant n}, (\mathcal{I}_{ij})_{1\leqslant i,j\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n})$ be a BIT model, then M is called *basic* if it satisfies the following two constraints for all $1 \leqslant i \neq j \leqslant n$ and $w \in W$,

(m1) for all $S \in \mathcal{T}_{ij}(w)$, if $\mathcal{B}_i \circ \mathcal{I}_{ij}(w) \subseteq S$, then $\mathcal{B}_i(w) \subseteq S$, where $\circ$ denotes the composition operator between two binary relations,[2]
(m2) $\mathcal{T}_{ij}(w) = \bigcap_{u\in\mathcal{B}_i(w)} \mathcal{T}_{ij}(u)$.

The class of basic BIT models is denoted by **BA**. The constraint (m2) essentially requires that an agent be self-aware of his/her trust towards other agents. This is a natural requirement for agents' mental attitudes because our agents should have some kind of introspective capability. On the other hand, (m1) makes a connection among the three classes of modal operators. It means that if an agent i believes that he has acquired the information φ from j and he trusts the judgement of j on the truth of φ, then he should

[2] $R_1 \circ R_2 = \{(x, y) \mid \exists z((x, z) \in R_1 \wedge (z, y) \in R_2)\}$.

1. Axioms:
   - P: all tautologies of the propositional calculus
   - B1: $[B_i\varphi \wedge B_i(\varphi \supset \psi)] \supset B_i\psi$
   - B2: $\neg B_i \bot$
   - B3: $B_i\varphi \supset B_i B_i\varphi$
   - B4: $\neg B_i\varphi \supset B_i \neg B_i\varphi$
   - I1: $[I_{ij}\varphi \wedge I_{ij}(\varphi \supset \psi)] \supset I_{ij}\psi$
   - I2: $\neg I_{ij}\bot$
   - C1: $B_i I_{ij}\varphi \wedge T_{ij}\varphi \supset B_i\varphi$
   - C2: $T_{ij}\varphi \equiv B_i T_{ij}\varphi$
2. Rules of Inference:
   - R1 (Modus ponens, MP): from $\vdash \varphi$ and $\vdash \varphi \supset \psi$ infer $\vdash \psi$
   - R2 (Generalization, Gen): from $\vdash \varphi$ infer $\vdash B_i\varphi$ and $\vdash I_{ij}\varphi$
   - R3: from $\vdash \varphi \equiv \psi$ infer $\vdash T_{ij}\varphi \equiv T_{ij}\psi$

Fig. 1. The axiomatic system BA for basic BIT.

assimilate the information into his belief base. These two constraints are represented by two natural axioms in our axiomatic system for basic BIT logic. The axiomatic system, called BA, is presented in Fig. 1.

The axioms B1–B4 correspond to the KD45 system for doxastic operators B_i. B1 means that the agents are perfect logical reasoners, so their beliefs are closed under logical consequence. B2–B4, corresponding to the serial, transitive and Euclidean properties of the $\mathcal{B}_i$ relation, stipulate respectively the consistency, positive introspection, and negative introspection of the agent's belief. The axioms I1 and I2 form the KD system for the information acquisition operators. Here, we assume that the operators describe not only the explicit information an agent directly acquires but also all consequences that are implicitly implied by it. Thus, if an agent acquires information φ, he also gets all logical consequence of φ at the same time. This is just what I1 asserts. Under the assumption, a source providing contradictory information will be useless, so we use axiom I2 to exclude the possibility that an agent can acquire contradictory information from a single source. However, note that this does not rule out the possibility that an agent can acquire contradictory information from multiple sources. Indeed, it is the notion of trust that can help to select what to believe when such a situation occurs. Finally, the connection axioms C1 and C2 correspond to the basic constraints (m1) and (m2) on the BIT models. C1 ties all three kinds of operators together and states when the acquired information should be assimilated into the beliefs, whereas C2 describes the mental states of an agent when he/she trusts the judgement of other agents. The Gen rule assures that valid wff is believed and acquired a prior, while R3 asserts that if an agent trusts another agent's judgement on some wff, then his trust is independent of the syntactic form of the wff.

Note that all axioms in this paper are interpreted as static constraints on the agent's belief, acquired information, and trust. If we interpret C1 dynamically, for example, as "if at time t_0, agent i believes that j has told him φ and i trusts j on the judgement of φ, then at time $t_0 + 1$, agent i will believe φ.", then this axiom does not naturally hold. If at time t_0, agent has believed $\neg\varphi$, then when receiving the information φ from j, he will face the dilemma of updating his belief or trust on j. He may decide not to trust j any more while

insisting on his belief in $\neg\varphi$ or give up his belief in $\neg\varphi$ while keeping his trust in j. The dynamic interpretation of the axiom only holds in the latter case. Nevertheless, no matter which way agent i has chosen, the final status should satisfy the axiom C1.

The derivability in the system is defined as follows. Let $\Sigma \cup \{\varphi\} \subseteq \Phi$, then φ is derivable from Σ in the system BA, written as $\Sigma \vdash_{\mathrm{BA}} \varphi$, if there is a finite sequence $\varphi_1, \ldots, \varphi_m$ such that every φ_i is an instance of an axiom schema in BA, a wff in Σ, or obtainable from earlier φ_j's by application of a rule in BA. When $\Sigma = \emptyset$, we simply write $\vdash_{\mathrm{BA}} \varphi$. The system BA is said to be sound if $\vdash_{\mathrm{BA}} \varphi$ implies $\models_{\mathbf{BA}} \varphi$ and complete if the converse holds.

Theorem 1. *The axiomatic system BA is sound and complete.*

Proof. The proof is based on the standard technique of canonical model construction in modal logic [10]. To make the paper self-contained, we include all proofs of this and the following theorems in Appendix A. □

3. Properties of trust

In the preceding section, we have described a set of basic axioms for BIT logic. In the system, we impose minimal constraints on the semantics of trust operators. However, there are still some useful theorems derivable in the system. For example, we have

$$\vdash_{\mathrm{BA}} B_i(I_{ij}\varphi \wedge I_{ik}\neg\varphi) \supset \neg(T_{ij}\varphi \wedge T_{ik}\neg\varphi) \tag{1}$$

and

$$\vdash_{\mathrm{BA}} \left[B_i(I_{ij}\varphi \wedge I_{ik}\neg\varphi) \wedge (T_{ij}\varphi \supset T_{ik}\neg\varphi)\right] \supset \neg T_{ij}\varphi. \tag{2}$$

The first says that if an agent acquired contradictory information from two sources, then not both sources are trusted by him, and the second further indicates that if one source is at least as trustworthy as the other, then the latter is not trusted. A more general form of (1) is the following derived rule:

$$\frac{\varphi_1 \wedge \cdots \wedge \varphi_m \supset \neg\varphi_{m+1}}{(B_i\varphi_{m+1} \wedge B_i(\bigwedge_{k=1}^{m} I_{ij_k}\varphi_k)) \supset \neg(\bigwedge_{k=1}^{m} T_{ij_k}\varphi_k)}. \tag{3}$$

This means that an agent does not trust all agents in a group if he believes they send him some information which is jointly incompatible with his belief.

On the other hand, there are some non-theorems of the system deserving further consideration. One notable example is if $\vdash_{\mathrm{BA}} \varphi \supset \psi$, could we infer $\vdash_{\mathrm{BA}} T_{ij}\varphi \supset T_{ij}\psi$? The intuition is that if we trust someone's judgement on a fact φ, should we also trust his judgement on a weaker fact ψ? At first sight, it seems tempting to have this as a derived rule of our system because, according to C1, when an agent acquired information φ, he will believe it due to the trust, so he will also believe the consequence ψ since he is a perfect reasoner. However, this does not mean that he will also accept the belief ψ if he is only informed of the fact ψ (less informative than φ). The situation can be illustrated by the following example.

Example 1. Let us consider a financial consultant j and a skeptical decision agent i, and φ and ψ denote respectively the facts: "The financial situation of company X is excellent." and "It is worthwhile to invest in company X." Then we may have $T_{ij}(\varphi \wedge (\varphi \supset \psi))$ because i believes that j has the capability to judge the financial situation of a company and the validity of the rules like $\varphi \supset \psi$. However, it is definitely not the case that i will believe that company X deserves his investment just because j tells him so without any justification, i.e., $B_i I_{ij}\psi \supset B_i\psi$ is not true. In this case $T_{ij}\psi$ should not hold, since otherwise we get a contradiction by C1.

This example also shows that $T_{ij}(\varphi \wedge \psi)$ does not imply $T_{ij}\varphi$ or $T_{ij}\psi$. Conversely, could we have $(T_{ij}\varphi \wedge T_{ij}\psi) \supset T_{ij}(\varphi \wedge \psi)$? The answer is also no because it is very likely that we have both $T_{ij}\varphi$ and $T_{ij}\neg\varphi$ at the same time, but we do not want to have $T_{ij}\bot$ as the result. When $\varphi \wedge \psi$ is logically consistent, it seems more appealing to have $(T_{ij}\varphi \wedge T_{ij}\psi) \supset T_{ij}(\varphi \wedge \psi)$ derivable from our system. In this case, we can add the following axiom

$$(T_{ij}\varphi \wedge T_{ij}\psi) \supset T_{ij}(\varphi \wedge \psi) \quad \text{if } \nvdash \neg(\varphi \wedge \psi).$$

However, since the axiom is subject to a non-derivability condition, it will make the system unnecessarily complex. In fact, since the belief operator satisfies the conjunction axiom $B_i\varphi \wedge B_i\psi \supset B_i(\varphi \wedge \psi)$, the above axiom for trust operator is rarely needed. Therefore, we do not stipulate such constraint in our logic. Instead, if the schema is true for some given φ and ψ, then the particular instance should be added as a premise of the agent specification.

Another derived rule shows that trust operators can play a role of filtering out noisy information. This rule is as follows:

$$\frac{\varphi \supset \psi}{B_i I_{ij}\varphi \wedge T_{ij}\psi \supset B_i\psi}. \tag{4}$$

In particular, we have

$$\vdash_{\mathrm{BA}} B_i I_{ij}(\varphi \wedge \psi) \wedge T_{ij}\psi \wedge B_i\neg\varphi \supset B_i\psi. \tag{5}$$

This means that even the whole piece of acquired information is contradictory with the agent's belief, he can still pick up some relevant part compatible with his belief.

Example 2. Let i denote a shopping agent who is in search of a digital camera and j a sale agent on behalf of an electronic product dealer. When j tells i that the digital camera of type AG007 is of high quality and low price, agent i may trust j that its price is indeed low but not that its quality is high. In this case, φ denotes "The price of type AG007 is low" and ψ "The quality of type AG007 is high", then we have both $B_i I_{ij}(\varphi \wedge \psi)$ and $T_{ij}\varphi$ but not $T_{ij}\psi$(nor $T_{ij}(\varphi \wedge \psi)$), so $B_i\varphi$ is derivable from the premises, though agent i does not necessarily believe both φ and ψ.

3.1. Symmetric trust and transferable trust

In the preceding discussion, we mentioned that it is very likely that sometimes both $T_{ij}\varphi$ and $T_{ij}\neg\varphi$ hold. Let us elaborate on this point further. This occurs when agent i trusts

the question-answering capability of j, so if i asks j whether fact φ holds, he is ready to accept either the positive or the negative answer j gives. This is particularly true when the agent is objective and neutral to the answer of any question. In artificial agent societies, this property of trust is especially useful, so we have defined a special system for it. A basic BIT model $M = (W, \pi, (\mathcal{B}_i)_{1\leqslant i\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n})$ is called symmetric if for all $w \in W$ and $1 \leqslant i \neq j \leqslant n$, it satisfies:

(m3) for all $S \subseteq W$ if $S \in \mathcal{T}_{ij}(w)$, then $\overline{S} \in \mathcal{T}_{ij}(w)$,

where $\overline{S} = W\backslash S$ is the complement of S with respect to W. The class of symmetric models is denoted by **SY** and the system SY will be the result of adding the following axiom to BA:

C3: $T_{ij}\varphi \supset T_{ij}\neg\varphi$.

The axiom C3 may not hold in the modeling of natural agents. For example, consider a critic j who is a very critical book reviewer. It is rarely the case that the critic says a book he reviews is good. Let φ denote the sentence "book X is very good". Then as a reader, agent i may trust j's judgement on φ being true but not the reverse, i.e., $T_{ij}\varphi \wedge \neg T_{ij}\neg\varphi$ holds for this case.

A special case of symmetric trust occurs when each agent specializes in different domain knowledge. For example, a medical agent specializes in health information, whereas a legal agent in law information, and so on. To model this kind of situation, let us assume $\Phi_1, \ldots, \Phi_n$ are pairwise disjoint subsets of Φ_0, and $\mathcal{L}(\Phi_i)$ is the set of BIT wffs formed only by atomic symbols in Φ_i for all $1 \leqslant i \leqslant n$. Then we can formulate a kind of trust, called *topical trust*, by the following nonstandard class of axioms:

$$T_{ij}\varphi \quad \text{if} \quad \nvdash \varphi, \nvdash \neg\varphi, \varphi \in \mathcal{L}(\Phi_j). \tag{6}$$

This class of axioms is nonstandard in at least two senses. First, a standard axiom schema can be instantiated by substituting any wffs into it, whereas the scope here is restricted to the subset $\mathcal{L}(\Phi_j)$ for each j. Second, the applicability of each axiom in the class depends on the non-derivability of φ and $\neg\varphi$ which is related to the whole axiomatic system including the axiom itself, so this makes the class of axioms not applicable in a constructive way and will complicate the reasoning unnecessarily. Furthermore, it seems also difficult to formulate a corresponding semantic constraint for the axioms, so we will not include them as logical axioms of our system. Instead, if necessary, for some subset of $\mathcal{L}(\Phi_j)$(for example, non-modal wffs), we can add $T_{ij}\varphi$ as the premises of reasoning for all φ in that subset.

Another property of trust deserving special attention is its transferability. Consider the following axiom:

C4: $B_i T_{jk}\varphi \supset T_{ik}\varphi$.

This means that if i believes that j trusts k, then i will also trust k due to the endorsement of j. This kind of trust will be called transferable trust. The system BA + C4 will be

denoted by TR. The corresponding constraint on the semantics may be easily formulated as follows:

(m4) $\bigcap_{u\in\mathcal{B}_i(w)} \mathcal{T}_{jk}(u) \subseteq \mathcal{T}_{ik}(w)$.

Let us call a basic model satisfying (m4) transferable model, and denote the class of all transferable models by **TR**. Then we have

Theorem 2. *Let L denote either SY or TR, then* $\vdash_{\mathrm{L}} \varphi$ *iff* $\models_{\mathbf{L}} \varphi$ *for any wff* φ.

A direct consequence of C1 and C4 is

$$B_i I_{ij} T_{jk}\varphi \wedge T_{ij} T_{jk}\varphi \supset T_{ik}\varphi. \tag{7}$$

This usually occurs in a recommendation system. When j recommends k to i by telling i that he trusts k about φ and i is familiar enough with j so that upon receiving the message, he knows that j is serious on the recommendation, then i would also trust k about φ due to the endorsement of j.

The notion of transferability must be carefully distinguished with that of transitivity. The latter is characterized by the following sentence

$$T_{ij}\varphi \wedge T_{jk}\varphi \supset T_{ik}\varphi. \tag{8}$$

A consequence of (8) is

$$T_{ij}\varphi \wedge T_{jk}\varphi \wedge B_i I_{ik}\varphi \supset B_i\varphi. \tag{9}$$

An argument against transitivity of trust in a form like (9) is given in [13]. The main argument is based on the fact that i can not know the communication between j and k, so he does not know whether k has told j about φ. Thus even if i believes $T_{jk}\varphi$, he has no legitimate basis to conclude $B_j\varphi$. Since the antecedent of (9) only contains $B_i I_{ik}\varphi$ instead of $B_i I_{ij}\varphi$, i has no sufficient reason to believe φ if he doubts whether k has lied to him (note that i does not trust k directly according to the antecedent of (9)). The argument in fact also denies the validity of the following sentence

$$T_{ij}\varphi \wedge B_i T_{jk}\varphi \wedge B_i I_{ik}\varphi \supset B_i\varphi, \tag{10}$$

which is derivable from the system TR.

In fact, what the argument reveals is that we should not expect that (10) holds universally for any kind of trust. This is indeed the case in our system since (10) does not follow from the basic system BA. However, it does not exclude the possibility that a special kind of trust can satisfy the transferability property. In particular, as mentioned above, when an agent would like to accept the recommendation of other agents, transferable trust would be useful in modeling the situation.

3.2. Cautious trust

If we analyze the factors of trust in detail, we can find the following two conditions are in general relevant for i to trust j on φ.

$$B_i(I_{ij}\varphi \supset B_j\varphi), \tag{11}$$

$$B_i(B_j\varphi \supset \varphi). \tag{12}$$

The first condition means that i believes that if j tells him φ then j himself believes φ, i.e., j is honest to him and the second means that i believes that if j believes φ, then φ in fact holds, i.e., j has good capability on evaluating the situation. Thus these two conditions correspond to two main factors of trust, i.e., the honesty and capability of the trusted agent. However, the two conditions are not necessary for an agent to commit himself to the trust because he cannot always be sure about the honesty and capability of the agent he would like to trust. For example, according to a past experience, he found that an agent was honest, however, he can not guarantee the agent will remain honest in the future. As for capability, any agent may make errors even if he has proved to be very capable in the past. Thus few agents would trust others only when they completely satisfy the two conditions. An agent will in general trust others if he has good confidence in their honesty and capability. For agents who insist on trusting others only when the two conditions are satisfied, we can call them cautious (or strict), and their trust is also called cautious (or strict). This is an ideal form of trust, so we can define a new class of modal operators T^c_{ij} as

$$T^c_{ij}\varphi =_{def} B_i\big[(I_{ij}\varphi \supset B_j\varphi) \wedge (B_j\varphi \supset \varphi)\big]. \tag{13}$$

Interestingly, the cautious trust also satisfies the axioms C1 and C2.

Proposition 1. *For any BIT wffs* φ

(1) $\vdash_{\mathrm{BA}} B_i I_{ij}\varphi \wedge T^c_{ij}\varphi \supset B_i\varphi$.
(2) $\vdash_{\mathrm{BA}} T^c_{ij}\varphi \equiv B_i T^c_{ij}\varphi$.

Proof. (1) Let ψ_1 and ψ_2 denote $I_{ij}\varphi$ and $B_j\varphi$ respectively, then, by substituting the definition of $T^c_{ij}\varphi$ into the wff, the theorem reduces to

$$\vdash_{\mathrm{BA}} B_i\psi_1 \wedge B_i\big((\psi_1 \supset \psi_2) \wedge (\psi_2 \supset \varphi)\big) \supset B_i\varphi$$

and this can be proved by P, B1, R1, and R2.

(2) This is a corollary of the more general theorem $\vdash_{\mathrm{BA}} B_i\psi \equiv B_i B_i\psi$ which can be proved by P, B1–B4 and R1. □

Since (13) is a very strict requirement for i to trust j on φ, it seems tempting to add the following axiom to our basic system for cautious agents. The system (BA + C5) will be denoted by CA.

C5: $T_{ij}\varphi \supset T^c_{ij}\varphi$.

Note that the axiom C5 is aimed at the modelling of cautious agents, so it does not hold for general agents. This is why we do not include it as an axiom of the system BA. To add the semantic constraint for C5, let us define a binary relation $\mathcal{T}^c_{ij} \subseteq W \times 2^W$ for every agent i and j such that $\mathcal{T}^c_{ij}(w, S)$ iff the following two conditions hold,

(i) for $u \in \mathcal{B}_i(w)$, $\mathcal{I}_{ij}(u) \subseteq S$ implies $\mathcal{B}_j(u) \subseteq S$, and
(ii) for $u \in \mathcal{B}_i(w)$, if $\mathcal{B}_j(u) \subseteq S$, then $u \in S$.

Consequently, a model is called cautious if, for each $w \in W$ and $1 \leqslant i \neq j \leqslant n$,

(m5) $\mathcal{T}_{ij}(w) \subseteq \mathcal{T}^c_{ij}(w)$

is satisfied.

Theorem 3. *Let* **CA** *denote the class of all cautious models, then* $\vdash_{\mathrm{CA}} \varphi$ *iff* $\models_{\mathbf{CA}} \varphi$ *for any wff* φ.

On the other hand, should the converse of C5 also hold since it seems reasonable for an agent to trust those with honesty and capability? Not necessarily because, if we add the converse of C5 to our system, an unintended side-effect $\vdash T_{ij}\bot$ will be derived since $\vdash_{\mathrm{BA}} T^c_{ij}\bot$ holds. It seems counterintuitive for an agent to trust another agent from this contradictory statement. This does not cause any problem in the agent's belief since we assume I_{ij} is a KD modal operator, so $T_{ij}\bot$ will not result in contradiction in the belief of agent i. Nevertheless, this seems to suggest that the converse of C5 does not necessarily hold for all rational trust. This is due to the fact that trust has some meaning of delegation in the sense that if agent i trusts agent j on φ, then he delegates the decision of the truth of φ to j.[3] Now, if agent i can believe the honesty of j about the truth of φ, and the capability of j in judging the truth of φ, then by a rational criteria, he can delegate the decision of the truth of φ to j. However, even in this case, agent i can still decide not to delegate the decision if he can decide the truth of φ by himself. In particular, when $\varphi = \bot$ or $\top$, agent i can decide the truth value of φ since he is a perfect reasoner in our logic. In this regard, the term "cautious trust" is somewhat abused to name the modal operator T^c_{ij} since it is not a special case of trust. More precisely, T^c_{ij} is only a kind of quasi-trust. This also shows that the derived operator T^c_{ij} is not adequate to formulate the notion of trust, so we have to use a primitive operator T_{ij} in our logic.

4. Properties of information acquisition

4.1. Ideal communication environment

In axiom C1, we use $B_i I_{ij}\varphi$ instead of $I_{ij}\varphi$ directly in the antecedent. This is due to a possibly insecure communication environment.

When i receives a message from j, if he can not exclude the possibility that someone pretending to be j has sent the message, then he does not necessarily believe that he has received the message from j. Thus we do not have $I_{ij}\varphi \supset B_i I_{ij}\varphi$. On the other hand, since someone pretending to be j may send a message to i and make i wrongly believe that he

[3] See [32] for a discussion of delegation logic in the context of computer system security.

indeed received the message from j, $B_i I_{ij}\varphi \supset I_{ij}\varphi$ does not necessarily hold, either. Now, if digital signature and secure communication is used, then when i receives some message with j's digital signature, he can believe this is indeed sent by j. When he believes j has sent him the message by recognizing the digital signature of j, it is impossible that it was counterfeit by another. Thus we have the following assumption under the ideal environment.

C6: $I_{ij}\varphi \equiv B_i I_{ij}\varphi$.

The corresponding semantic constraint for C6 is:

(m6) $\mathcal{B}_i \circ \mathcal{I}_{ij} = \mathcal{I}_{ij}$.

A basic model satisfying (m6) will be called an ideal communication model, and the class of such models is denoted by **IC**. The system IC is the result of adding C6 to BA and replacing C1 with

C1′: $I_{ij}\varphi \wedge T_{ij}\varphi \supset B_i\varphi$.

Theorem 4. *For any wff* φ, $\vdash_{\mathrm{IC}} \varphi$ *iff* $\models_{\mathbf{IC}} \varphi$.

4.2. Logic of utterance

In an ideal communication environment, if private communication is allowed, then it is possible that $I_{ij}\varphi$ and $I_{kj}\neg\varphi$ hold at the same time. That is, agent j may tell one agent the truth but lie to another. However, if the only communication channel among the agents is a public one, i.e., an agent can tell others something only by announcing it in public, then we can add both the axiom $I_{ij}\varphi \equiv I_{kj}\varphi$ and the semantic constraint that for all $1 \leqslant i \neq j \neq k \leqslant n$, $\mathcal{I}_{ij} = \mathcal{I}_{kj}$ and still have soundness and completeness results. However, in this case, we can even further simplify the language of the BIT logic. For each j, the class of operators $I_{1j}, \ldots, I_{nj}$ can be replaced by a single operator U_j. The meaning of $U_i\varphi$ is then "the agent i utters φ". This is a logic of belief, utterance and trust (BUT). The formation rules, semantics, and axiomatic system of BUT logic are obtained by replacing I_{ij} by U_j uniformly in those of BIT logic. The resultant axiomatic system is named BU and is listed in Appendix A (see Figs. A.3 and A.4). Let CU6 denote the axiom $U_j\varphi \equiv B_i U_j\varphi$ and IU denote BU + CU6. Let (mu1) and (mu6) denote the results of replacing $\mathcal{I}_{ij}$ by $\mathcal{U}_j$ in (m1) and (m6) respectively and let **BU** (respectively **IU**) denote the classes of BUT models satisfying (mu1) and (m2) (respectively (mu1), (m2) and (mu6)). Then we have

Theorem 5. *Let L denote either BU or IU, then* $\vdash_{\mathrm{L}} \varphi$ *iff* $\models_{\mathbf{L}} \varphi$ *for any BUT wffs* φ.

A logic for utterance and knowledge in the single-agent case has been proposed in [44] for the analysis of the well-known liar paradox, where the epistemic operator is an S5 modal operator, and the utterance operator is a KD45 operator and an axiom like CU6 holds therein. Though the system (called KU there) is different from ours, it is similar to

our IU to some extent, so we can also define a liar in IU or BU. Formally, an agent i is called an *intentional liar* if $U_i\varphi \wedge B_i\neg\varphi$ is true, and an *irresponsible liar* if $U_i\varphi \wedge \neg B_i\varphi$ is true. Obviously, an intentional liar is also an irresponsible one.

In the context of IU, an agent i is said to be *honest*[4] if he is not a irresponsible liar (i.e., $U_i\varphi \supset B_i\varphi$ for all φ of BUT logic), and *frank* if $B_i\varphi \supset U_i\varphi$ for all φ of BUT logic. An extreme case where all agents are honest and frank may occur when all agents inform others of their total belief. In this case, the operators U_i can be further removed from the BUT logic and we can get a logic of belief and trust (BT). In the basic BT system (by replacing all U_i by B_i), we can prove the theorem

$$B_j\varphi \wedge T_{ij}\varphi \supset B_i\varphi. \tag{14}$$

This means that if i trust j, then i will believe what j believes. If there is a mutual trust between i and j, i.e., $T_{ij}\varphi \wedge T_{ji}\varphi$, then the belief of i and j is equivalent. The system BT is useful in modeling a set of cooperative agents in which each agent has unlimited access to other agents' knowledge base. Then the beliefs of each agent would be directly influenced by the beliefs of other agents according to his trust in them.

5. Related works

5.1. Trust in multi-agent systems

In [8], it is argued that trust is a notion of crucial importance for multi-agent systems. While the authors regard trust as both a mental state, and a social attitude and relation, we consider specifically the influence of trust on the assimilation of acquired information into an agent's belief. By using modal logic tools, we semantically and axiomatically characterize the relationship among belief, information acquisition and trust. Among the existing works on the application of the notion of trust to multi-agent systems, the one most related to ours is [15]. In [15], trust is considered to be an attitude of an agent who believes that another agent has a given property, so unlike in our definition, trust is analyzed as a derived concept instead of a primitive one. The context of [15] is the same as ours. There is a set of information sources, called agents, which can communicate with each other. Thus there are three classes of modal operators for the agents. The first is for belief, denoted by B_i, which is a KD normal modal operator.

The second is for strong belief, denoted by K_i, which is not only a KD normal modal operator but also satisfied by the following axiom (KT)

$$K_i(K_i\varphi \supset \varphi).$$

The relationship between belief and strong belief is characterized by the axiom (KB)

$$K_i\varphi \supset B_i\varphi.$$

[4] However, since we consider belief instead of knowledge, an honest agent may still make errors, so it is possible $U_i\varphi \wedge \neg\varphi$ holds for an honest agent i.

Since our belief operator B_i is a KD45 one, it also satisfies (KT). consequently, it plays both roles of belief and strong belief operators. In other words, we do not distinguish belief and strong belief in our logic. However, our belief operators satisfy more properties than those in [15] (whether strong or not). In particular, it has positive and negative introspection properties which are commonly accepted as the characterization of perfect reasoners.

The third class of modal operators are for information action, where $I_{i,j}\varphi$ denotes that agent i has informed agent j about φ, so syntactically, there is a reverse in the direction of information flow from our information acquisition operators (i.e., $I_{ij}\varphi = I_{j,i}\varphi$).[5] However, putting aside the tiny difference in syntax, the main difference between information action operators and information acquisition ones is that the former are minimal modal operators instead of normal ones. Thus the only property for $I_{i,j}$ in [15] is the inference rule (RE).

(1) If $\vdash \varphi \equiv \psi$, then $\vdash I_{i,j}\varphi \equiv I_{i,j}\psi$.

Thus the information action operator will only consider the explicit information communicated by some agent to another one. This is not enough for us since sometimes one agent explicitly trusts only part of the information which is communicated to him by another. An illustrative situation is given in Example 2 where both $B_i I_{ij}(\varphi \wedge \psi)$ and $T_{ij}\varphi$ hold, but not $T_{ij}(\varphi \wedge \psi)$, so if I_{ij} is not a normal modal operator, we could not derive $B_i\varphi$ by axiom C1. Furthermore, it is assumed that the information action operator $I_{i,j}$ satisfies axioms

(OBS1) $I_{i,j}\varphi \supset K_j I_{i,j}\varphi$,
(OBS2) $\neg I_{i,j}\varphi \supset K_j \neg I_{i,j}\varphi$,

which are collectively equivalent to

$$K_j I_{i,j}\varphi \equiv I_{i,j}\varphi \equiv \neg K_j \neg I_{i,j}\varphi.$$

The former is simply equivalent to our axiom C6 in system IC for an ideal communication environment. Though the latter is also intuitively reasonable for an ideal communication environment, it is not necessary for our aim of eliminating the modality B_i from the left of the axiom C1. However, the axiom (OBS2) is important in the reasoning of trust on completeness which is lacking in our framework.

Based on the basic modalities of belief, strong belief, and information action, different sorts of trust are defined in [15]. These include sincerity, credibility, cooperativity, vigilance, validity, and completeness.

$$Tsinc_{i,j}(\varphi) =_{def} K_i(I_{j,i}\varphi \supset B_j\varphi),$$
$$Tcred_{i,j}(\varphi) =_{def} K_i(B_j\varphi \supset \varphi),$$
$$Tvigi_{i,j}(\varphi) =_{def} K_i(\varphi \supset B_j\varphi),$$
$$Tcoop_{i,j}(\varphi) =_{def} K_i(B_j\varphi \supset I_{j,i}\varphi),$$

[5] Note the comma between j and i.

$$Tval_{i,j}(\varphi) =_{def} Tsinc_{i,j}(\varphi) \wedge Tcred_{i,j}(\varphi),$$
$$Tcomp_{i,j}(\varphi) =_{def} Tvigi_{i,j}(\varphi) \wedge Tcoop_{i,j}(\varphi).$$

Regarding the two-fold roles played by our belief operators, the definitions of $Tsinc_{i,j}(\varphi)$ and $Tcred_{i,j}(\varphi)$ correspond exactly to Eqs. (11) and (12) (i.e., the honesty and capability factors of a trust), so trust in the validity is just what we define as cautious trust T^c_{ij}. In proposition 1, it has been shown that cautious trust satisfies axioms C1 and C2. Analogously, a counterpart of C1, $Tval_{i,j}(\varphi) \wedge K_i I_{j,i}\varphi \supset K_i\varphi$ also holds in the logic of [15]. However, since K_i does not satisfy axioms like B3 and B4, the axiom C2 does not have a counterpart for the notions of trust defined in [15]. This means that agents are not necessarily self-aware of their trust.

Though the analysis of trust as a derived concept provides insightful understanding of the notions, this will sometimes result in counter-intuitive consequences. For example, by the normality of K_i, (OBS2), and the definition of trust on validity, we can derive $Tsinc_{i,j}(\varphi)$ from $\neg I_{j,i}\varphi$. In other words, i trusts j for his sincerity on φ just because j has not informed him of φ, though in practice, the definitions are only used in the forward reasoning in the examples presented in [15]. This problem does not arise in our system because we consider trust as a primitive concept and add appropriate axioms for characterizing its relationship with some derived ones (e.g., C5) when necessary. Furthermore, the primitive notion of trust may also accommodate some subjective factors of trust, such as emotion.

Fairly speaking, the problem mentioned above is partly due to the use of material implication in the definitions of different notions of trust. Thus, when generalizing them to the graded ones, it is replaced by a conditional operator. For example, the trust level of i about the sincerity of j on φ being α is defined as

$$Tsinc^{\alpha}_{i,j}(\varphi) =_{def} K_i(I_{j,i}\varphi \Rightarrow_{\alpha} B_j\varphi),$$

where $\Rightarrow_{\alpha}$ is a binary connective in conditional logic [10]. Though the feature of graded trust is completely absent in our logic, we think it would be more useful when combined with information fusion operators [33]. It deserves further investigation to extend our logic along this direction. (Also see Section 6 for further research in information fusion.)

When these varied types of trust are interpreted in a specific context of a database, they may be used in expressing some properties of a database, such as informational validity and completeness [7]. Let s and db be two agents denoting the system administrator and the database, then $Tcred_{s,db}(\varphi)$ and $Tvigi_{s,db}(\varphi)$ correspond respectively to "s knows db is valid for φ" and "s knows db is complete for φ". Nevertheless, the notions of trust are also applicable to reasoning about the safety of information in a database. We will discuss an application along this direction in Section 5.3.

5.2. Trust in agent communication language

The notions of trust have also been extensively used in Agent Communication Language (ACL). ACL is important for multi-agent systems since the coordination and collaboration between agents depends on effective inter-agent communication [9]. So far, two main ACLs are KQML (Knowledge Query and Manipulation Language) and

FIPA-ACL (Foundation for Intelligent Physical Agents). The semantics for KQML is given in [31], whereas that for FIPA-ACL is based on a quantified multi-modal logic originally proposed in [14]. Both semantics depend on the speech act theory [38]. The main primitives of an ACL are called performatives which are interpreted as communicative acts. A communicative act can change the mental state of the receiver of a message just as an ordinary act can change the physical state of the outside world. Therefore, the semantics of each performative is specified by its precondition and effect.

Much effort has been spent on the semantics of ACLs [5,25,30,42,48] and, some of the most important performatives, such as "INFORM" and "REQUEST", have been precisely defined. In particular, the performative INFORM is closely related to our information acquisition operator. According to [25,30], the effect of INFORM is to construct mutual belief between two agents under the sincerity and competence assumption of the message sender. Formally, under the assumption that the sender agent x is sincere, the following result holds

$$\models \big(\text{DONE (INFORM } x\ y\ e\ p\ t)\big) \wedge \big(\text{MB } x\ y\ \big((\text{BEL } x\ p) \supset p\big)\big) \Rightarrow (\text{MB } x\ y\ p), \quad (15)$$

where $\Rightarrow$ denotes a defeasible implication. By omitting the temporal and event parameters t and e, the formula (DONE (INFORM $x\ y\ e\ p\ t$)) is roughly equivalent to $I_{yx}p$ in our logic, whereas the formula (BEL $x\ p$) is exactly $B_x p$ in our logic. Though the mutual belief operator MB is given by a fixed point definition, it is shown that (MB $x\ y\ p$) can be thought of as an infinite conjunction of (BEL $x\ p$), (BEL $y\ p$), (BEL x (BEL $y\ p$)), ... and so on. Therefore, the first difference between INFORM act and our information acquisition operator is that the former is aim to construct mutual belief between two agents which is stronger than the receiver's belief in the transmitted information.

While mutual belief is needed for coordination and collaboration, it is impossible in some contexts. In particular, if an agent posts a message in a news group or mailing list, he does not know who will receive it. Therefore, if the receiving agent does not reply directly to the sending agent, it is impossible to construct mutual belief. However, it is still possible that the receiving agent's mental state can be changed by the message. Our characteristic axiom C1 can be applied in such situations since it is stated from the viewpoint of the receiving agent.

There are even arguments that mutual belief can never be established via message transmission if the communication channel is not fully reliable [20,40,41]. However, the difficulty can be overcome by use of default [28,30,36]. This is why the defeasible implication $\Rightarrow$ is used above. Since axiom C1 is simply a static constraint on the belief of the receiving agent, we do not need the default assumption, and the ordinary implication can be used.

The second difference between the notion of trust used in ACL and that defined in our framework is that the former is defined by the sincerity and competence of the sending agent instead of a primitive notion. For example, in [25], trust is explicitly defined as

$$(\text{TRUST } i\ j\ p) \equiv \big(\text{BEL } i\ (\text{BEL } j\ p)\big) \supset (\text{BEL } i\ p)$$

which is weaker than the condition (12) for cautious trust. As argued in Section 3.2, it is inadequate in some cases to treat trust as a derived notion from the belief on the sincerity and competence of the sending agent.

Finally, though trust is an important notion in the semantics of ACL, the interpretation of ACL performatives and protocols needs a far more general theory of agency. While some mental attitudes, such as intention and commitment, have been extensively studied in ACL research, more research is needed to incorporate these notions into our logic.

5.3. Logics for information safety in database

An earlier attempt on the reasoning of information safety in database context has been made in [12,16], resulting in the development of three logical systems S, S′, and S″. In the context of database management, there are some agents, called "information sources", which store messages in a database *DB*. These messages can be read by another agent called "system", who knows the meaning of every stored message. There is yet another special agent, called DB administrator, who has the meta-information about the reliability of the source agents. Two kinds of beliefs are distinguished by the system agent: *ordinary belief* is the information incorporated into the database by any agents, whereas *true belief* is only that inserted by reliable agents. To model this situation, a modal logic S is first proposed in [16] based on the signaling act theory [26]. Then for the purpose of computational implementation, two simplified versions of S, called S' and S" are proposed in [12]. Since S has the complete features of the sequence of logics, we will focus our comparison to it. The logic includes the modalities E_i, B_s, B_i, K_s, and K_{adm} for information source agents i, the system agent s and the administrator agent *adm*. The modality E_i is called an action operator, so $E_i p$ means intuitively "agent i brings it about that p". The language of the logic is rather fine-grained in the sense that it explicitly distinguishes the form and meaning of a message. However, by using an auto-naming convention, a wff $E_i(in.DB('\varphi'))$ is used to denote that agent i inserts a piece of information φ into *DB*. The wff then corresponds to our information acquisition operator $I_{si}\varphi$ (or $U_i\varphi$ since s is the only information receiver in the context of database) which means that agent i sends information φ to the system agent. The modalities B_s and B_i are exactly the belief operators in our language, and the knowledge (or true belief) operator K_s is defined as $K_s\varphi = B_s\varphi \wedge \varphi$. Furthermore, K_{adm} is the knowledge operator for the administrator agent.

In addition to the axioms for the action operators, four main axioms of logic S are as follows:

- (OBS): $E_i p \supset K_s E_i p$.
- (S): $K_s(E_i(in.DB('\varphi')) \supset B_i\varphi)$.
- (BEL)[6] : $K_s B_i\varphi \supset K_s B_s\varphi$.
- (SAF): $K_{adm}(E_i p \supset q) \supset K_s(E_i p \supset q)$.

Furthermore, the definition of safety operator is $\text{safe}(i,\varphi) =_{def} K_{adm}(E_i(in.DB('\varphi')) \supset \varphi)$.

[6] In the presentation of S in [12], the axiom is given as $K_s B_i\varphi \supset K_s B\varphi$, however, since $B\varphi$ denote the ordinary belief of the system, it should be equivalent to $B_s\varphi$.

In the axiom (OBS), if we substitute $in.DB('\varphi')$ into p, then by using the correspondence between $E_i(in.DB('\varphi'))$ and $I_{si}\varphi$ and the definition of $K_s\varphi$, we can derive

$$I_{si}\varphi \supset K_s I_{si}\varphi, \tag{16}$$

which implies

$$I_{si}\varphi \supset B_s I_{si}\varphi. \tag{17}$$

This is just an instance of the forward implication of our axiom C6. On the other hand, though the stronger property $K_s I_{si}\varphi \supset I_{si}\varphi$ holds trivially in S, it does not possess the reverse implication of C6, i.e., $B_s I_{si}\varphi \supset I_{si}\varphi$.

By our interpretation, the axiom (S) can be rewritten as

$$K_s(I_{si}\varphi \supset B_i\varphi), \tag{18}$$

which means that the system agent knows each agent i is honest. The (BEL) axiom says that if s knows an agent i believes φ, then s knows he also believes it himself. By combining (16), (18) and (BEL), we can derive

$$I_{si}\varphi \supset B_s\varphi. \tag{19}$$

Thus the system agent will believe every piece of information he receives from any source agents, no matter whether they are trustworthy. Since it is required that all belief operators are KD45 ones in S, the source agents can not send jointly contradictory information to the system. This is not compatible with the situation we would like to model in multi-agent communication, though it is possible to impose such constraint in the database management context.

The wff (19) is stronger than C1′ since it drops the condition $T_{si}\varphi$ from the left side of the implication. Thus the ordinary belief of the system agent is only a superset of the information he received from any sources, and does not rely on the reliability of the information sources. How the trust operator can play a role is in the formation of truth belief. If we substitute p and q in axiom (SAF) with $in.DB('\varphi')$ and φ respectively, then by the definition of safe(i,φ), we can derive

$$\text{safe}(i,\varphi) \supset (B_s I_{si}\varphi \supset B_s\varphi), \tag{20}$$

which is an instance of our C1 with safe(i,φ) corresponding to $T_{si}\varphi$. However, the definition of safe(i,φ) in fact says more than this wff since it assures $I_{si}\varphi \supset \varphi$, i.e., the judgement of i on φ is fully trustworthy, by the fact that $K_s p = B_s p \wedge p$. Since the axiom (S) has asserted the honesty of agent i, safe(i,φ) is actually closer to our cautious trust $T^c_{si}\varphi$.

Based on the comparison, we find that many important notions of the system S can also be captured by our systems. This demonstrates the potential of applying our framework to the database context. Furthermore, our framework provides more detailed analysis for the properties of trust operators, so it is more appropriate for modeling of agent communication and belief formation.

5.4. Trust in computer security

The notions of trust have played an important role in computer security. However, they were usually used in an intuitive sense and not formally defined until the 1980s. One of the earliest attempts to formalize the notions of trust in a modal logic framework is made in [37]. In that formalization, the logic of belief is taken as the basis, and trust is defined as a proper axiom added to the logic. Thus, in this logic, there are no operators for trust and information acquisition as our logic has and trust is a derived notion in the framework. For example, authenticity trust between agents i and j about the key of agent k is defined as

$$B_i B_j B_k \,\mathrm{owner}(key_k, k) \supset B_i B_k \,\mathrm{owner}(key_k, k),$$

which is derivable from $T_{ij}\,\mathrm{owner}(key_k, k)$ in the system BT by R2, C2, B3, B4, and (14). This shows that, even with strong assumptions (i.e., ideal communication environment and the honesty and frankness of the agents), our logic can still derive some useful notions of trust. Though the basic logic in [37] is very simple, a method to map a formal trust specification to mechanisms for its implementation in distributed systems has been also developed.

However, to analyze the notions of trust, the logic of belief is obviously inadequate since it does not distinguish what an agent believes and what he says. Thus, a more expressive logic is needed for finer analysis. One of the most important logics for the purpose is the logic of authentication developed by Burrows, Abadi, and Needham [6], called BAN logic. The logic has been subsequently refined and extended in some further works [1,43]. In BAN logic, there are two operators closely related to our information acquisition and trust operators. These are the **said** and **controls** operators. For the purpose of comparison, let us translate them into the unary modal notations S and C, so

$$S_i\varphi = P_i \quad \textbf{said} \quad \varphi,$$
$$C_i\varphi = P_i \quad \textbf{controls} \quad \varphi,$$

where P_i is the agent i, usually called *principal* in computer security literature. The rule connecting **said**, **controls**, and **believes** operators is called the *jurisdiction rule* and is formulated as

$$B_i C_j \varphi \wedge B_i B_j \varphi \supset B_i \varphi \tag{21}$$

in [6] and as

$$C_j \varphi \wedge S_j \varphi \supset \varphi \tag{22}$$

in [43].[7] By applying rule R2, we can derive from (22) the following

$$B_i C_j \varphi \wedge B_i S_j \varphi \supset B_i \varphi. \tag{23}$$

In BAN logic, it is usually assumed that principals are honest, so $S_j\varphi \supset B_j\varphi$ holds and (23) can also be derived from (21). It can be seen that (23) has some analogy with our axiom C1 (or CU1). Indeed, (23) can be viewed as a common intersection of the two variants of the

[7] Note that we ignore the temporal parameters from the original formulation for simplicity.

jurisdiction rule and our axiom C1 (and CU1) if we intuitively interpret $B_i C_j \varphi$ as "agent i trusts j about φ", and $B_i S_j \varphi$ as "i believes that j told him φ".

In spite of the apparent analogy between axiom C1 (and CU1) and (23), we can not overlook the conceptual difference between our logic and BAN logic. First, in BAN logic, the **control** operator means the objective jurisdiction of an agent as described in (22), so if $T_{ij}\varphi$ means that an agent i regards another agent j as an authority on φ, then it is roughly equivalent to $B_i C_j \varphi$ in BAN logic. However, it has been also argued in [13] that trust can not be identified with jurisdiction in all cases, so we need a primitive operator for the representation of trust. Second, since the **said** operator in BAN logic is mainly for transmitting messages for authentication, it can be assumed that principals are always honest. However, in multi-agent systems which we would like to model, one agent may cheat another for the interest of himself, so the honesty assumption does not automatically hold in our logic. Third, **said** is in fact more like the utterance operator U_i in our system BU since it only specifies the sender of the message. Thus, when it is used together with the belief operator as in the wff $B_i S_j \varphi$, it is implicitly assumed that the receiver of the **said** operator is agent i. However, in [13], it is shown that sometimes the implicit assumption is the source of some counterintuitive derivation, so the explicit specification of the receiver is necessary. This is why our information acquisition operator is indexed by both the sending and receiving agents. Finally, from a technical viewpoint, BAN logic is comparatively more complicated (though also more expressive in some aspects) than ours. No completeness proof of its axiomatic system is given, whereas our logic mainly concentrates on the basic features of the trust, belief, and information acquisition operators, thus making complete axiomatization possible. Furthermore, our axiomatic systems are rather modular in the sense that further properties of trust can be added to the existing systems due to our choice of neighborhood semantics for the trust operator.

5.5. *A logic of delegation*

It is mentioned above that trust has some sense of delegation. Recently, a delegation logic [32] has been proposed for reasoning about the authorization decision in distributed systems. In that logic (called D1LP), two kinds of statements are basic. The first, called direct statement, is of the form

$$X \texttt{ says } p$$

where X is a principal which roughly corresponds to an agent in our logic, and p is an atomic formula in first-order logic. The second, called delegation statement, has the form

$$X \texttt{ delegates } p\hat{}d \texttt{ to } PS$$

where X and p is as above, d is a positive integer or the asterisk symbol '$*$' representing the delegation depth, and PS is a relatively complicated structure called principal structure which can be viewed as a group of agents semantically. Roughly speaking, direct statements correspond to belief modality, whereas delegation statements correspond to trust in our logic. However, since there is no information acquisition modality in D1LP, the interpretation of delegation statement is in fact closer to Eq. (14). Moreover, the common restriction of both statements is that p must be an atomic formula in classical logic (or

a literal when it is extended to D2LP, a delegation logic for handling negation) while we allow any wffs in the scope of our modal operators. Though in the context of computer security, this seems enough, we indeed have to represent more complex forms of formulas such as nested modality in the information gathering context. On the other hand, D1LP allows a more complicated structure for agents and agent groups. For example, agent i may partly trust j_1 and j_2 individually but fully trust them jointly, so if j_1 or j_2 (but not both) tell him φ individually, he will not entertain the belief. However, if they both tell him the fact, then he will accept it. This seems to suggest a direction for further extension of our logic by considering the notions of group trust. For example, let G be a subset of $\{1, 2, \ldots, n\}$ and $k \leqslant |G|$ a natural number, then the trust operators can be generalized to the form of $T_{iG|k}$ for any G, k and $i \notin G$. The intuitive meaning of $T_{iG|k}\varphi$ is that agent i will delegate his decision of the truth of φ to any k agents in G, so we have the generalization of C1:

$$\bigvee_{G'}\left(B_i \bigwedge_{j\in G'} I_{ij}\varphi\right) \wedge T_{iG|k}\varphi \supset B_i\varphi,$$

where G' ranges over all k-element subsets of G.

6. Future work

Though we have discussed the notion of trust extensively, many interesting problems still remain untouched in the presentation above. In this section, we briefly remark on some of these possible research directions.

First, the dual parts of sincerity, credibility, and validity, i.e., respectively, cooperativity, vigilance, and completeness defined in [15] are totally lacking in our logic. Research in this area may be very useful in the reasoning of closed-world databases. For example, if j is a database agent monitoring the train timetable, i is a client agent who believes that j would inform him the time of all departure trains, and i does not inform j of a train departure at 3:50, then j will believe that there is not one. This dual kind of trust may need axioms such as

$$T_{ij}\varphi \wedge B_i\varphi \supset B_i I_{ij}\varphi.$$

Therefore, it is worthwhile to investigate how our framework can be extended with these dual notions of trust.

Second, as mentioned in Section 5.1, a research direction related to graded trust is information fusion in which an agent must decide what to accept among possibly inconsistent information from different sources with various degrees of reliability. In multi-source reasoning [11], it is shown how literal information can be merged, and it is suggested that for information of general form, the belief revision approach of Katsuno and Mendelzon [29] can be used. In [33], multi-source reasoning logic is extended with the distributed knowledge operator of multi-agent epistemic logic [20] so that the belief of different agents can be merged according to their degrees of reliability. However, since information acquisition operators are not included in that logic, it can not model the notion of agent communication. On the other hand, in BIT logic, trust is only a qualitative

notion, so it will be interesting to generalize it quantitatively or ordinally so that we can merge acquired information from sources with various degrees of reliability. Therefore, the integration of these two logics will produce one with richer expressive power. Some preliminary work along this direction has been done in [18].

Third, in Section 3.1, a special case of symmetry trust, called topical trust, is considered without standard axiomatization. This problem may be remedied by introducing the topics of propositions into the language. For example, in a logic for aboutness [17], a sorted binary predicate $A(t, `p')$ is used to denote "sentence '$p$' is about topic $t$". If our BIT language is extended with such a predicate, then we can formulate axioms as: $A(t, `\varphi') \supset T_{ij}\varphi$ when j is specialized at topic t, or more strongly, as: $(A(t_1, `\varphi') \vee \cdots \vee A(t_k, `\varphi')) \equiv T_{ij}\varphi$ when the set of topics at which j is specialized are $\{t_1, \ldots, t_k\}$. However, further research is needed to see how the semantics can be changed to accommodate this syntactic extension.

Forth, we will consider the dynamics of information acquisition. So far, the I_{ij} operators only describe the static fact that some information is acquired. However, we can also consider how information acquisition action causes the transition of the belief state. There are essentially two way to do this. One is to index the modal operators with a time stamp so that $\Box^t\varphi$ means that $\Box\varphi$ holds in time t for any modal operator $\Box = B_i, T_{ij}$, or I_{ij}. Then we can have a dynamic counterpart of C1 as follows:

$$B_i^t I_{ij}^t\varphi \wedge T_{ij}^t\varphi \wedge \neg B_i^t\neg\varphi \supset B_i^{t+1}\varphi.$$

The other way is to use the dynamic logic framework [23]. Let $[I_{ij}\varphi]$ denote a dynamic operator for each $1 \leqslant i \neq j \leqslant n$ and wff φ. We can try to develop an update semantics for these operators [47] along the direction of the works reported in [22,45,46]. Then the above axiom is rewritten as

$$T_{ij}\varphi \wedge \neg B_i\neg\varphi \supset [I_{ij}\varphi]B_i\varphi.$$

Fifth, though axiomatic systems provide an elegant characterization of the notions we want to model, some more realistic proof methods, such as tableau methods, Gentzen sequent calculus, or resolution methods, must be developed for the purpose of implementation. In this regard, the approaches adopted in [3,4,21] are ideal starting points.

Last, though we mainly consider the influence of trust on the acceptance of acquired information as belief, on the reverse, we can also try to induce the trust degree of an agent according to how much information acquired from him has been accepted as belief in the past. To do this, we must first add the temporal dimension to the semantics of our logic. Then the trust degree of i on j at time t, denoted by $d_{ij}^t : W \to [0, 1]$ can be defined by

$$d_{ij}^t(w) = \frac{|\{\varphi : w, t-1 \models B_i\varphi \wedge I_{ij}\varphi\}|}{|\{\varphi : w, t-1 \models I_{ij}\varphi\}|}.$$

This formula can be seen as a realization of the trust update function introduced in [27] where the experience values used in [27] are concretely computed as the proportion of information accepted by the receiving agent. However, according to the current semantics of I_{ij}, it is an implicit information acquisition operator, so $\{\varphi : w, t-1 \models I_{ij}\varphi\}$ is in general infinite. Thus, to make the definition meaningful, we should only consider the explicit information acquired by i from j. This means that we will change the semantics of I_{ij} to a minimal one, and require that $\mathcal{I}_{ij}(w)$ is finite for any $w \in W$. In this way, it is

expected to model the quite complicated phenomenon of multi-agent communication with different trust degrees in a logical system.

7. Conclusion

In this paper, we characterize the notions of trust, belief, and information acquisition for multi-agent systems in a modal logic framework. Some different properties of trust and information acquisition operators are discussed and axiomatized, showing the usefulness of logical tools in formulating the properties of multi-agent systems. The proposed framework is also compared with related works in computer security and database management, demonstrating the potential applications of our logic in such fields.

The logic framework presented in this paper can guide the implementation of multi-agent systems in the following way. The logic provides a rigorous semantics for the notions of trust, information acquisition, and belief, and characterizes the relationship among these notions precisely. Therefore, it is possible to verify whether an agent system actually complies with the semantics if the states of the system can be appropriately connected with the mental states of agents. A general framework for ascribing knowledge (or belief) in multi-agent systems has been proposed in [20]. The basic principle is to construct a Kripke model by the inter-agent communication history. The information acquisition operator can be easily interpreted in such a system if each agent can remember the information he has received. Also, since we adopt a minimal semantics for the trust operator, it can be interpreted in such a system if each agent i has a knowledge base KB_i such that each element of KB_i is of the form (j, φ) which means that $T_{ij}\varphi$ holds. In this way, each agent can reason about the mental states of himself and other agents, so that a multi-agent system complying with the semantics provided in this paper can be implemented by the knowledge-based program proposed in [20]. Though this is only a rough guideline for the design of multi-agent systems based on our semantics, it is expected that detail implementation will further confirm the principle.

Acknowledgements

We would like to thank two anonymous referees for their helpful comments.

Appendix A. Proof of theorems

The theorems to be proved in this appendix all concern soundness and completeness of logical systems. For the ease of reference, we list the above-mentioned model constraints and axioms in Figs. A.1–A.4 and include their correspondence with the system names in Table A.1. In Table A.1, we add two core systems L_0 and LU_0 as the origin of our axiomatic systems, where L_0 is just the BA system without the connection axioms C1 and C2, and LU_0 is the result of replacing I_{ij} in L_0 by U_j. Thus all our systems are the result of adding some connection axioms in Fig. A.4 to these two core systems.

Table A.1
The correspondence between axioms and model constraints

Modalities	System	Axioms	Constraints	Completeness
B_i, T_{ij}, I_{ij}	L_0	P, B1–B4, I1–I2, R1–R3	none	*
	BA	L_0 + C1–C2	m1–m2	Thm 1
	SY	L_0 + C1–C3	m1–m3	Thm 2
	TR	L_0 + C1–C2, C4	m1–m2, m4	Thm 2
	CA	L_0 + C1–C2, C5	m1–m2, m5	Thm 3
	IC	L_0 + C1′, C2, C6	m1–m2, m6	Thm 4
B_i, T_{ij}, U_i	LU_0	replace I_{ij} by U_j in L_0	none	*
	BU	LU_0 + CU1, C2	mu1, m2	Thm 5
	IU	LU_0 + CU1′, C2, CU6	mu1, m2, mu6	Thm 5

(m1) for all $S \in \mathcal{T}_{ij}(w)$, if $\mathcal{B}_i \circ \mathcal{I}_{ij}(w) \subseteq S$, then $\mathcal{B}_i(w) \subseteq S$.
(m2) $\mathcal{T}_{ij}(w) = \bigcap_{u \in \mathcal{B}_i(w)} \mathcal{T}_{ij}(u)$.
(m3) for all $S \subseteq W$ if $S \in \mathcal{T}_{ij}(w)$, then $\overline{S} \in \mathcal{T}_{ij}(w)$.
(m4) $\bigcap_{u \in \mathcal{B}_i(w)} \mathcal{T}_{jk}(u) \subseteq \mathcal{T}_{ik}(w)$.
(m5) $\mathcal{T}_{ij}(w) \subseteq \mathcal{T}^c_{ij}(w)$.
(m6) $\mathcal{B}_i \circ \mathcal{I}_{ij} = \mathcal{I}_{ij}$.
(mu1) for all $S \in \mathcal{T}_{ij}(w)$, if $\mathcal{B}_i \circ \mathcal{U}_j(w) \subseteq S$, then $\mathcal{B}_i(w) \subseteq S$.
(mu6) $\mathcal{B}_i \circ \mathcal{U}_j = \mathcal{U}_j$.

Fig. A.1. The model constraints.

1. Axioms:
   - P: all tautologies of the propositional calculus
   - B1: $[B_i\varphi \wedge B_i(\varphi \supset \psi)] \supset B_i\psi$
   - B2: $\neg B_i \bot$
   - B3: $B_i\varphi \supset B_i B_i\varphi$
   - B4: $\neg B_i\varphi \supset B_i \neg B_i\varphi$
   - I1: $[I_{ij}\varphi \wedge I_{ij}(\varphi \supset \psi)] \supset I_{ij}\psi$
   - I2: $\neg I_{ij} \bot$
2. Rules of Inference:
   - R1 (Modus ponens, MP): from $\vdash \varphi$ and $\vdash \varphi \supset \psi$ infer $\vdash \psi$
   - R2 (Generalization, Gen): from $\vdash \varphi$ infer $\vdash B_i\varphi$ and $\vdash I_{ij}\varphi$
   - R3: from $\vdash \varphi \equiv \psi$ infer $\vdash T_{ij}\varphi \equiv T_{ij}\psi$

Fig. A.2. The axiomatic system L_0.

As usual, the verification of soundness is a routine checking of the validity of the axioms and the validity-preservation of the inference rules in respective systems, so it is left up to the readers to check. For the completeness part, Let L denote one of our logical systems and **L** denote its corresponding class of models. A wff φ is L-inconsistent if its negation $\neg\varphi$ can be proved in L. Otherwise, φ is L-consistent. A set Σ of wffs is said to be L-inconsistent if there is a finite subset $\{\varphi_1, \ldots, \varphi_k\} \subseteq \Sigma$ such that the wff $\varphi_1 \wedge \cdots \wedge \varphi_k$ is L-inconsistent;

1. Axioms:
   - P: all tautologies of the propositional calculus
   - B1: $[B_i\varphi \wedge B_i(\varphi \supset \psi)] \supset B_i\psi$
   - B2: $\neg B_i \bot$
   - B3: $B_i\varphi \supset B_i B_i\varphi$
   - B4: $\neg B_i\varphi \supset B_i \neg B_i\varphi$
   - U1: $[U_i\varphi \wedge U_i(\varphi \supset \psi)] \supset U_i\psi$
   - U2: $\neg U_i \bot$
2. Rules of Inference:
   - R1 (Modus ponens, MP): from $\vdash \varphi$ and $\vdash \varphi \supset \psi$ infer $\vdash \psi$
   - RU2 (Generalization, Gen): from $\vdash \varphi$ infer $\vdash B_i\varphi$ and $\vdash U_i\varphi$
   - R3: from $\vdash \varphi \equiv \psi$ infer $\vdash T_{ij}\varphi \equiv T_{ij}\psi$

Fig. A.3. The axiomatic system LU_0.

- C1: $B_i I_{ij}\varphi \wedge T_{ij}\varphi \supset B_i\varphi$.
- C2: $T_{ij}\varphi \equiv B_i T_{ij}\varphi$.
- C3: $T_{ij}\varphi \supset T_{ij}\neg\varphi$.
- C4: $B_i T_{jk}\varphi \supset T_{ik}\varphi$.
- C5: $T_{ij}\varphi \supset T^c_{ij}\varphi$.
- C6: $I_{ij}\varphi \equiv B_i I_{ij}\varphi$.
- C1′: $I_{ij}\varphi \wedge T_{ij}\varphi \supset B_i\varphi$.
- CU1: $B_i U_j\varphi \wedge T_{ij}\varphi \supset B_i\varphi$.
- CU1′: $U_j\varphi \wedge T_{ij}\varphi \supset B_i\varphi$.
- CU6: $U_j\varphi \equiv B_i U_j\varphi$.

Fig. A.4. Additional axioms.

otherwise, Σ is L-consistent. A maximal L-consistent set of wffs (L-MCS) is a consistent set χ of wffs such that whenever ψ is a wff not in χ, then $\chi \cup \{\psi\}$ is L-inconsistent.

On the other hand, φ is L-satisfiable iff there exists a model M in **L** and a possible world w such that $w \models_M \varphi$, otherwise φ is L-unsatisfiable. Sometimes the prefix L will be omitted without confusion. To prove the completeness, we will show that every L-consistent wff is **L**-satisfiable. To prove the result, we use the standard canonical model construction technique in modal logic [10]. Let us first consider the case where L is an extension of L_0. The case for LU_0-extended systems can be proved in an analogous way.

A canonical L-model $M^* = (W, \pi, (\mathcal{B}_i)_{1\leqslant i\leqslant n}, (\mathcal{I}_{ij})_{1\leqslant i\neq j\leqslant n}, (\mathcal{T}_{ij})_{1\leqslant i\neq j\leqslant n})$ is such that

- $W = \{w_\chi \mid \chi \text{ is an L-MCS}\}$, in other words, each possible world corresponds precisely to an L-MCS.
- $\pi : \Phi_0 \to 2^W$ is defined by $\pi(p) = \{w_\chi \mid p \in \chi\}$.
- $\mathcal{B}_i(w_{\chi_1}, w_{\chi_2})$ iff $\chi_1/B_i \subseteq \chi_2$, where $\chi_1/B_i = \{\varphi \mid B_i\varphi \in \chi_1\}$.
- $\mathcal{I}_{ij}(w_{\chi_1}, w_{\chi_2})$ iff $\chi_1/I_{ij} \subseteq \chi_2$, where $\chi_1/I_{ij} = \{\varphi \mid I_{ij}\varphi \in \chi_1\}$.
- $\mathcal{T}_{ij}(w_\chi) = \{[\varphi] \mid T_{ij}\varphi \in \chi\}$, where $[\varphi] = \{w_{\chi'} \mid \varphi \in \chi'\}$.

Note that by this construction, we can show that $[\varphi] = [\psi]$ implies $\vdash_{\mathrm{L}} \varphi \equiv \psi$. If $\nvdash_{\mathrm{L}} \varphi \equiv \psi$, then either $\varphi \wedge \neg\psi$ or $\neg\varphi \wedge \psi$ is L-consistent. In either case, we can have an L-MCS containing one of φ or ψ but not the other, so $[\varphi] \neq [\psi]$. Therefore, we can have $[\varphi] \in \mathcal{T}_{ij}(w_\chi)$ iff $T_{ij}\varphi \in \chi$ which will be used implicitly in the proof of some lemmas below.

In what follows, we will first show that a canonical L-model is indeed a model in $\mathbf{L}_0$.

Lemma A.1.

(1) *Each $\mathcal{B}_i$ in a canonical L-model is serial, transitive, and Euclidean.*
(2) *Each $\mathcal{I}_{ij}$ in a canonical L-model is serial.*

Proof. These results can be easily proved by the axioms B2–B4 and I2. We show the transitivity of $\mathcal{B}_i$ as an example. If both $\mathcal{B}_i(w_{\chi_1}, w_{\chi_2})$ and $\mathcal{B}_i(w_{\chi_2}, w_{\chi_3})$ hold, then for any $B_i\varphi \in \chi_1$, we have $B_iB_i\varphi \in \chi_1$ by axiom B3, so $B_i\varphi \in \chi_2$ and $\varphi \in \chi_3$ by the definition of $\mathcal{B}_i$. Thus $\chi_1/B_i \subseteq \chi_3$, and so $\mathcal{B}_i(w_{\chi_1}, w_{\chi_3})$ holds. □

The most important result for such construction is the truth lemma.

Lemma A.2 (Truth lemma). *For any wff φ and L-MCS χ, we have $w_\chi \models_{M^*} \varphi$ iff $\varphi \in \chi$.*

Proof. By induction on the structure of the wff, the only interesting case is the wff of the form $\Box\psi$ for some modality $\Box = B_i, I_{ij}$ or T_{ij}. Let us take $\Box = B_i$ and T_{ij} as examples.

First, for $\Box = B_i$, by definition, $w_\chi \models_{M^*} B_i\psi$ iff for all $w_{\chi'} \in \mathcal{B}_i(w_\chi)$, $w_{\chi'} \models_{M^*} \psi$ iff for all $\chi/B_i \subseteq \chi'$, $\psi \in \chi'$ (by induction hypothesis) iff $\chi/B_i \cup \{\neg\psi\}$ is L-inconsistent iff $B_i\psi \in \chi$ when B_i is a normal modal operator [10]. However, by the axioms P and B1, rules MP and Gen, B_i is indeed a normal modal operator.

Second, for $\Box = T_{ij}$, by definition of satisfaction, $w_\chi \models_{M^*} T_{ij}\psi$ iff $|\psi| \in \mathcal{T}_{ij}(w_\chi)$ iff $[\psi] \in \mathcal{T}_{ij}(w_\chi)$ (by induction hypothesis) iff $T_{ij}\psi \in \chi$ by the definition of canonical models. □

By combining these two lemmas, it has sufficed to prove that L_0 is complete. For if $\nvdash_{\mathrm{L}_0} \varphi$, then $\neg\varphi$ is L_0-consistent, so we can find an L_0-MCS containing $\neg\varphi$ and consequently, φ is not valid in the canonical model, i.e., $\not\models_{\mathbf{L}_0} \varphi$. To prove that L is complete for L = BA, SY, TR, CA, IC, we must show that the respective canonical model is in the corresponding model class.

A.1. Proof of Theorem 1

Lemma A.3. *The canonical BA-model M^* is in* **BA**.

Proof. We will verify that the constraints (m1) and (m2) are satisfied.

(1) (m1): if $[\varphi] \in \mathcal{T}_{ij}(w_\chi)$ and $\mathcal{B}_i \circ \mathcal{I}_{ij}(w_\chi) \subseteq [\varphi]$, then by definition, $T_{ij}\varphi \in \chi$ and φ is in every BA-MCS containing $\Sigma =_{def} \{\psi \mid B_i I_{ij}\psi \in \chi\}$. From the latter, it follows that there exist $\psi_1, \ldots, \psi_k \in \Sigma$ such that

$$\vdash_{\mathrm{BA}} (\psi_1 \wedge \cdots \wedge \psi_k) \supset \varphi.$$

Thus by axioms P, B1, I1, and rules R1 and R2, it follows that $B_i I_{ij}\varphi \in \chi$. Then by the axiom C1, $B_i\varphi \in \chi$ also holds. This means that for any $w_{\chi'} \in \mathcal{B}_i(w_\chi)$, $\varphi \in \chi'$, i.e, $w_{\chi'} \in [\varphi]$, so $\mathcal{B}_i(w_\chi) \subseteq [\varphi]$.

(2) (m2): $[\varphi] \in \mathcal{T}_{ij}(w_\chi)$ iff $T_{ij}\varphi \in \chi$ iff $B_i T_{ij}\varphi \in \chi$ (by C2) iff $T_{ij}\varphi \in \chi'$ for any χ' such that $w_{\chi'} \in \mathcal{B}_i(w_\chi)$ iff $[\varphi] \in \bigcap_{u \in \mathcal{B}_i(w_\chi)} \mathcal{T}_{ij}(u)$. □

Then Theorem 1 follows directly from Lemmas A.1–A.3.

A.2. Proof of Theorem 2

Lemma A.4.

(1) *The canonical SY-model M^* is in* **SY**.
(2) *The canonical TR-model M^* is in* **TR**.

Proof.

(1) To prove that M^* satisfies (m3), assume that $[\varphi] \in \mathcal{T}_{ij}(w_\chi)$, then $T_{ij}\varphi \in \chi$ and so $T_{ij}\neg\varphi \in \chi$ by C3. Thus $\overline{[\varphi]} = [\neg\varphi] \in \mathcal{T}_{ij}(w_\chi)$.
(2) To prove that M^* satisfies (m4), assume that $[\varphi] \in \bigcap_{w_{\chi'} \in \mathcal{B}_i(w_\chi)} \mathcal{T}_{jk}(w_{\chi'})$, then $T_{jk}\varphi \in \chi'$ for any χ' such that $\chi/B_i \subseteq \chi'$ by definition of $\mathcal{B}_i$ and $\mathcal{T}_{jk}$, so $T_{jk}\varphi \in \chi/B_i$ and $B_i T_{jk}\varphi \in \chi$. Therefore, by axiom C4, it can be derived that $T_{ik}\varphi \in \chi$, so $[\varphi] \in \mathcal{T}_{ik}(w_\chi)$. □

Then Theorem 2 follows directly from Lemmas A.1, A.2, and A.4.

A.3. Proof of Theorem 3

Lemma A.5. *The canonical CA-model M^* is in* **CA**.

Proof. To verify that the model satisfies (m5), we assume $[\varphi] \in \mathcal{T}_{ij}(w_\chi)$, then $T_{ij}\varphi \in \chi$, so by axiom C5 and the definition of T^c_{ij}, we have $I_{ij}\varphi \supset B_j\varphi \in \chi'$ and $B_j\varphi \supset \varphi \in \chi'$ for any $w_{\chi'} \in \mathcal{B}_i(w_\chi)$. From the former, it follows that if $\mathcal{I}_{ij}(w_{\chi'}) \subseteq [\varphi]$, then $\mathcal{B}_j(w_{\chi'}) \subseteq [\varphi]$ and from the latter, it follows that if $\mathcal{B}_j(w_{\chi'}) \subseteq [\varphi]$, then $w_{\chi'} \in [\varphi]$ for $w_{\chi'} \in \mathcal{B}_i(w_\chi)$, so $[\varphi]$ satisfies the two conditions for $\mathcal{T}^c_{ij}(w_\chi, [\varphi])$, i.e., $[\varphi] \in \mathcal{T}^c_{ij}(w_\chi)$. □

Then Theorem 3 follows directly from Lemmas A.1, A.2, and A.5.

A.4. Proof of Theorem 4

Lemma A.6. *The canonical IC-model M^* is in* **IC**.

Proof. To verify that the model satisfies (m6), we need only note that $\mathcal{B}_i \circ \mathcal{I}_{ij}(w_{\chi_1}, w_{\chi_2})$ iff $\chi_1/B_i I_{ij} \subseteq \chi_2$, where $\chi_1/B_i I_{ij} = \{\varphi \mid B_i I_{ij}\varphi \in \chi_1\}$. Then by axiom C6, we have $\chi_1/B_i I_{ij} = \chi_1/I_{ij}$, so the result follows immediately. For the constraints (m1), since the axiom C1' is equivalent to C1 according to axiom C6, so the result proved in Lemma A.3 still holds. □

A.5. Proof of Theorem 5

To prove the theorem, the canonical model constructed in the beginning of the appendix must be modified slightly. We merely have to replace $\mathcal{I}_{ij}$ with $\mathcal{U}_i$ with the following definition:

- $\mathcal{U}_i(w_{\chi_1}, w_{\chi_2})$ iff $\chi_1/U_i \subseteq \chi_2$, where $\chi_1/U_i = \{\varphi \mid U_i\varphi \in \chi_1\}$.

Then Lemma A.1 (with each $\mathcal{U}_i$ being serial) and Lemma A.2 also hold for the canonical LU_0-model. Since BU and IU are extensions of LU_0, we must only prove the following lemma to finish the proof of Theorem 5.

Lemma A.7.

(1) *The canonical BU-model M^* is in* **BU**.
(2) *The canonical IU-model M^* is in* **IU**.

Proof. The proof that the canonical BU-model satisfies (mu1) and (m2) is analogous to that for Lemma A.3, and that the canonical IU-model satisfies (mu1), (m2), and (mu6) is analogous to that for Lemma A.6. □

References

[1] M. Abadi, M. Burrows, B. Lampson, G. Plotkin, A calculus for access control in distributed systems, ACM Trans. Programming Language Syst. 15 (4) (1993) 706–734.
[2] L. Åqvist, Deontic logic, in: D.M. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, Vol II: Extensions of Classical Logic, D. Reidel, 1984, pp. 605–714.
[3] M. Baldoni, Normal multimodal logics: Automatic deduction and logic programming extension, PhD Thesis, Dipartimento di Informatica, Universitá degli Studi di Torino, 1998.
[4] M. Baldoni, L. Giordano, A. Martelli, A tableau calculus for multimodal logics and some (un)decidability results, in: H. de Swart (Ed.), Proc. of the International Conference on Analytic Tableaux and Related Methods, in: Lecture Notes in Artificial Intelligence, Vol. 1397, Springer, Berlin, 1998, pp. 44–59.
[5] P. Bretier, D. Sadek, A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction, in: J.P. Müller, M. Wooldridge, N.R. Jennings (Eds.), Intelligent Agents III, in: Lecture Notes in Artificial Intelligence, Vol. 1193, Springer, Berlin, 1997, pp. 189–204.

[6] M. Burrows, M. Abadi, R. Needham, A logic of authentication, ACM Trans. Comput. Syst. 8 (1) (1990) 18–36.
[7] J. Carmo, R. Demolombe, A.J.I. Jones, An application of deontic logic to information system constraints, Fundamenta Informaticae 48 (2–3) (2001) 165–181.
[8] C. Castelfranchi, R. Falcone, Principle of trust for MAS: Cognitive anatomy, social importance, and quantification, in: Y. Demazeau (Ed.), Proc. of the 3rd International Conference on Multi Agent Systems, Paris, 1998, pp. 72–79.
[9] B. Chaib-Draa, F. Dignum, Trends in agent communication language, Comput. Intelligence 18 (2) (2002) 89–101.
[10] B.F. Chellas, Modal Logic: An Introduction, Cambridge University Press, Cambridge, 1980.
[11] L. Cholvy, A logical approach to multi-sources reasoning, in: M. Masuch, L. Pólos (Eds.), Knowledge Representation and Reasoning under Uncertainty, in: Lecture Notes in Artificial Intelligence, Vol. 808, Springer, Berlin, 1994, pp. 183–196.
[12] L. Cholvy, R. Demolombe, A. Jones, Reasoning about the safety of information: From logical formalization to operational definition, in: Z.W. Ras, M. Zemankova (Eds.), Methodologies for Intelligent Systems: 8th International Symposium, in: Lecture Notes in Artificial Intelligence, Vol. 869, Springer, Berlin, 1994, pp. 488–499.
[13] B. Christianson, W.S. Harbison, Why isn't trust transitive?, in: M. Lomas (Ed.), Security Protocols: International Workshop, in: Lecture Notes in Artificial Intelligence, Vol. 1189, Springer, Berlin, 1997, pp. 171–176.
[14] P.R. Cohen, H.J. Levesque, Communicative actions for artificial agents, in: V. Lesser (Ed.), Proc. of the 1st International Conference on Multi Agent Systems, San Francisco, CA, MIT Press, Cambridge, MA, 1995, pp. 65–72.
[15] R. Demolombe, To trust information sources: A proposal for a modal logic framework, in: C. Castelfranchi, Y.H. Tan (Eds.), Trust and Deception in Virtual Societies, Kluwer Academic, Dordrecht, 2001.
[16] R. Demolombe, A. Jones, Deriving answers to safety queries, in: R. Demolombe, T. Imielinski (Eds.), Nonstandard Queries and Nonstandard Answers, Oxford University Press, Oxford, 1994, pp. 113–129.
[17] R. Demolombe, A.J.I. Jones, On sentences of the kind "Sentence 'P' is about topic T", in: H.J. Ohlbach, U. Reyle (Eds.), Logic, Language and Reasoning—Essays in Honour of Dov Gabbay, Kluwer Academic, Dordrecht, 1999, pp. 115–133.
[18] R. Demolombe, C.J. Liau, A logic of graded trust and belief fusion, in: Proc. of the 4th Workshop on Deception, Fraud and Trust in Agent Societies, 2001, pp. 13–25.
[19] D.C. Dennett, The Intentional Stance, MIT Press, Cambridge, MA, 1987.
[20] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Reasoning about Knowledge, MIT Press, Cambridge, MA, 1996.
[21] M.C. Fitting, Proof Methods for Modal and Intuitionistic Logics, D. Reidel, Dordrecht, 1983.
[22] J. Gerbrandy, W. Groeneveld, Reasoning about information change, J. Logic Language Inform. 6 (1997) 147–169.
[23] D. Harel, Dynamic logic, in: D.M. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logic, Vol II: Extensions of Classical Logic, D. Reidel, Dordrecht, 1984, pp. 497–604.
[24] J. Hintikka, Knowledge and Belief, Cornell University Press, Ithaca, NY, 1962.
[25] M.J. Huber, S. Kumar, P.R. Cohen, D.R. McGee, A formal semantics for proxycommunicative acts, in: J.-J. Meyer, M. Tambe (Eds.), Intelligent Agents VIII: Agent Theories, Architectures, and Languages, in: Lecture Notes in Artificial Intelligence, Vol. 2333, Springer, Berlin, 2002, pp. 221–234.
[26] A.J.I. Jones, Towards a formal theory of communication and speech acts, in: Intentions in Communication, MIT Press, Cambridge, MA, 1990, pp. 141–160.
[27] C.M. Jonker, J. Treur, Formal analysis of models for the dynamics of trust based on experiences, in: F.J. Garijo, M. Boman (Eds.), Multi-Agent System Engineering: Proc. of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World, in: Lecture Notes in Artificial Intelligence, Vol. 1647, Springer, Berlin, 1999, pp. 221–232.
[28] Y. Katagiri, Belief coordination by default, in: Proc. of the 2nd International Conference on Multi Agent Systems, MIT Press, Cambridge, MA, 1996, pp. 142–149.
[29] H. Katsuno, A. Mendelzon, Propositional knowledge base revision and minimal change, Artificial Intelligence 52 (1991) 263–294.

[30] S. Kumar, M.J. Huber, P.R. Cohen, D.R. McGee, Towards a formalism for conversation protocols using joint intention theory, Comput. Intelligence 18 (2) (2002) 174–228.
[31] Y. Labrou, T. Finin, Semantics and conversations for an agent communication language, in: Proc. of IJCAI-97, Nagoya, Japan, 1997.
[32] N. Li, B.N. Grosof, J. Feigenbaum, A logic-based knowledge representation for authorization with delegation, in: Proc. of 12th IEEE Computer Security Foundations Workshop, IEEE Press, 1999, pp. 162–174.
[33] C.J. Liau, A conservative approach to distributed belief fusion, in: Proc. of the Third International Conference on Information Fusion, 2000, pp. MoD4–1.
[34] C.J. Liau, Logical systems for reasoning about multi-agent belief, information acquisition and trust, in: Proc. of the 14th European Conference on Artificial Intelligence, Berlin, IOS Press, Amsterdam, 2000, pp. 368–372.
[35] J.-J.Ch. Meyer, W. van der Hoek, Epistemic Logic for AI and Computer Science, Cambridge University Press, Cambridge, 1995.
[36] C.R. Perrault, An application of default logic to speech act theory, in: P. Cohen, J. Morgan, M. Pollack (Eds.), Intentions in Communication, MIT Press, Cambridge, MA, 1990, pp. 161–185.
[37] P.V. Rangan, An axiomatic basis of trust in distributed systems, in: Proc. of 1988 IEEE Symposium on Security and Privacy, IEEE Press, 1988, pp. 204–211.
[38] J.R. Searle, Speech Acts, Cambridge University Press, Cambridge, 1969.
[39] Y. Shoham, Agent-oriented programming, Artificial Intelligence 60 (1) (1993) 51–92.
[40] M.P. Singh, The intentions of teams: Team structure, endodeixis, and exodeixis, in: Proc. of 13th European Conference on Artificial Intelligence, Brighton, Wiley, New York, 1998, pp. 303–307.
[41] M.P. Singh, A social semantics for agent communication languages, in: F. Dignum, M. Greaves (Eds.), Issues in Agent Communication, in: Lecture Notes in Artificial Intelligence, Vol. 1916, Springer, Berlin, 2000, pp. 31–45.
[42] I.A. Smith, P.R. Cohen, Towards a semantics for an agent communications language based on speech-acts, in: Proc. of AAAI-96, Portland, OR, 1996, pp. 24–31.
[43] S. Stubblebine, R. Wright, An authentication logic supporting synchronization, revocation, and recency, in: Proc. of 3rd ACM Conference on Computer and Communications Security, ACM, New York, 1996, pp. 95–105.
[44] A. Tzouvaras, Logic of knowledge and utterance and the liar, J. Philos. Logic 27 (1998) 85–108.
[45] B. van Linder, W. van der Hoek, J.-J.Ch. Meyer, Tests as epistemic updates, in: A. Cohen (Ed.), Proc. of ECAI-94, Amsterdam, Wiley, New York, 1994, pp. 331–335.
[46] B. van Linder, W. van der Hoek, J.-J.Ch. Meyer, Actions that make you change your mind, in: I. Wachsmuth, C. Rollinger, W. Brauer (Eds.), Proc. of KI-95, in: Lecture Notes in Artificial Intelligence, Vol. 981, Springer, Berlin, 1995, pp. 185–196.
[47] F. Veltman, Defaults in update semantics, J. Philos. Logic 25 (1996) 221–261.
[48] M. Wooldridge, Verifiable semantics for agent communication language, in: Y. Demazeau (Ed.), Proc. of 3rd International Conference on Multi Agent Systems, Paris, 1998, pp. 349–356.
[49] M. Wooldridge, N. Jennings, Intelligent agents: Theory and practice, Knowledge Engineering Review 10 (2) (1995) 115–152.