Logical Systems for Reasoning about Multi-agent Belief, Information Acquisition and Trust

Churn-Jung Liau[1]

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 tools, we characterize the relationship among belief, information acquisition and trust both semantically and axiomatically. The belief and information acquisition are respectively represented by KD45 and KD normal modal operators, whereas trust is expressed 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.

1 Introduction

Recently, more and more software agents have been designed to solve the information search problem arising from rapid growth of internet information . The agents can search through the web and try to find and filter out information matching the user's need. However, not all internet information sources are reliable. Some web sites are out-of-date, some news provide wrong information, and someone even intentionally spreads rumor or deceives by anonymity. From the viewpoint of agent societies, each agent plays both the roles of information provider and receiver, so the information search process can be seen as the communication between two agents and a receiver has to decide whether he can believe the received information according to his trusting attitude toward the provider.

In [10], an agent is characterized by mental attitudes, such as knowledge, belief, obligation, and commitment. This view of agent, in accordance with the intentional stance proposed in [4], has been widely accepted as a convenient way for the analysis and description of complex systems[15]. The model of these attitudes has been the traditional concern of philosophical logic. Some logics derived from the philosophical analysis have been applied to the modeling of AI and distributed systems[9, 5]. 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[2]. Following the approach, we would like to propose a doxastic logic with modalities for representing the trusting attitudes and the information transmission between agents, and then discuss how one agent's belief is influenced by the others based on his trust toward other agents and the information he acquires. 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 agent j's judgement 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 meeting the above-mentioned requirement. The syntax, semantics, and a basic axiomatic system of the logic will be presented. Then, some additional assumptions will be considered to produce variants of the basic logic. 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[2]. 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 $1 \leq i, j \leq n$, and
- if φ and ψ are wffs, then $\varphi \vee \psi$ is, too.

As usual, other classical Boolean connectives can be defined as abbreviations.

The possible-worlds semantics provides a general framework for the modeling of knowledge and belief[5]. 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 with 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 component. Since the assessment of credibility of an agent can only depend on his past records, we can not guarantee 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.

[1] Institute of Information Science, Academia Sinica, Taipei, 115, Taiwan, email: liaucj@iis.sinica.edu.tw

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[2]. Formally, a BIT model is a tuple $(W, \pi, (\mathcal{B}_i)_{1\leq i\leq n}, (\mathcal{I}_{ij})_{1\leq i,j\leq n}, (\mathcal{T}_{ij})_{1\leq i,j\leq n})$, 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[2] 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 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 that 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 be a BIT model as above 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 $w \in \pi(p)$ for every $p \in \Phi_0$,
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 $|\varphi| = \{u \in W : u \models \varphi\}$ 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 $\forall\psi \in \Sigma, w \models_M \psi$ 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 the communication with others. This is definitely not what we want to model. Though we do not want an agent to believe blindly what the other agents tell him, it is indeed inevitably his belief should be influenced by the information he acquired from the agents he trusts. Based on the consideration, we will impose some constraints on the BIT models. Let M be a BIT model as above, then M is called *basic* if it satisfies the following two constraints for all $1 \leq i, j \leq n$ 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 $\mathcal{B}_i(w) \subseteq S$,
(m2) $\mathcal{T}_{ij}(w) = \bigcap_{u\in\mathcal{B}_i(w)} \mathcal{T}_{ij}(u)$.

[2] 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 $\forall w, u, v(\mathcal{R}(w, u) \wedge \mathcal{R}(w, v) \Rightarrow \mathcal{R}(u, v))$.

The class of basic BIT models is denoted by **BA**. The constraint (m2) essentially requires that an agent trust the judgement of other agents iff he believe it is the case. This is a natural requirement for agents' mental attitudes. On the other hand, (m1) make 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 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.

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: $B_i\varphi \supset B_iB_i\varphi$

   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: $B_iI_{ij}\varphi \wedge T_{ij}\varphi \supset B_i\varphi$

   C2: $T_{ij}\varphi \equiv B_iT_{ij}\varphi$

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 $\vdash T_{ij}\varphi \equiv T_{ij}\psi$

Figure 1. The axiomatic system BA for basic BIT

The axioms B1-B4 correspond to the KD45 system for doxastic operator B_i, B1 means that the agents are perfect logical reasoners, so their belief 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 acquires directly but also all consequences that implicitly implied by it, so if an agent acquires the information φ, he also get 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 that the notion of trust can help to select what to believe when such 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 belief, whereas C2 describes the mental states of an agent when he trust on 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 trust another agent's judgement on some wff, then his trust is independent of the syntactic form of the wff.

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.*

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

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

The first says that if an agent acquired contradictory information from two sources, then not both sources are reliable, and the second further indicates that for the two contradictory sources, the less reliable one is distrusted. Since a source can only be trusted or distrusted and no intermediary degree is allowed, a source which is contradictory with the more reliable ones must be distrusted. A more general form of (1) is the following derived rule:

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

This means that an agent can not trust all agents in a group if he believes they have sent him some pieces of information which are 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 the first sight, it seems tempting to have this as a theorem 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 fact "The financial situation of company X is excellent." and "It is worthwhile to invest on company X." Then we may have $T_{ij}(\varphi \wedge (\varphi \supset \psi))$ because i considers 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 company X deserves his investment just because j tell him so without any justification, i.e., $B_iI_{ij}\psi \supset B_i\psi$ is not true, so $T_{ij}\psi$ does not hold due to 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 negative 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.

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

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

In particular, we have

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

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 an information search agent who has known that there is a flight from New York to San Francisco on Thursday and want to know the fare of the ticket. Suppose j is a web site which can provide such information but due to system error, it display that the flight is on Friday as well as the correct fare of the ticket. Then if the search agent trusts j's fare information, he can still use the information even the total information displayed in the site is inconsistent with his belief. Note that we do not explain why i should trust j on the fare information even j contains some wrong information. This may be due to the past experience of i and need some learning mechanism to establish it. Here we are just interested in what will happen if i has trusted j on the fare information.

3.1 Symmetry 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 the agent i trust on the question-answering capability of j, so if i asks j whether the fact φ holds, he is ready to accept either the positive or the negative answer once j gives it. This is particularly true when the agent is objective and neutral to the answer of any question. In the artificial agent societies, this property of trust is especially useful, so we define a special system for it. A basic BIT model $M = (W, \pi, (\mathcal{B}_i)_{1\leq i\leq n}, (\mathcal{I}_{ij})_{1\leq i,j\leq n}, (\mathcal{T}_{ij})_{1\leq i,j\leq n})$ is called symmetry if for all $w \in W$ and $1 \leq i, j \leq n$, 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 symmetry 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 that is a very critical book reviewer. It is rarely the case that the critic said that a book reviewed by him is good. Let φ denote the sentence "the book X is very good". Then as a reader, the 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 symmetry trust occurs when each agent is specialized in some different domain knowledge. For example, a medical agent is specialized 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 pairwisely disjoint subsets of Φ_0 and $\mathcal{L}(\Phi_i)$ be the set of BIT wffs formed only by atomic symbols in Φ_i for all $1 \leq i \leq n$. Then we can formulate a kind of trust, called *topical trust*, by the following nonstandard axiom:

$$T_{ij}\varphi \quad \text{if} \quad \not\vdash \varphi, \not\vdash \neg\varphi, \varphi \in \mathcal{L}(\Phi_j) \tag{6}$$

This axiom 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 the axiom 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 axiom not applicable in a constructive way and will result in the non-monotonicity of the system. Furthermore, it seems also difficult to formulate a corresponding semantic constraint for the axiom, so we will not include it as a logic axiom 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: $B_iT_{jk}\varphi \wedge T_{ij}T_{jk}\varphi \supset T_{ik}\varphi$.

This means that if i trusts the evaluation of j on the reliability of k and he believes that j indeed trust 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) for any $S \subseteq W$, if $\mathcal{B}_i(w) \subseteq \mathcal{T}_{jk}^{-1}(S) \in \mathcal{T}_{ij}(w)$, then $S \in \mathcal{T}_{ik}(w)$,

where $\mathcal{T}_{jk}^{-1}(S) = \{v \in W \mid (v, S) \in \mathcal{T}_{jk}\}$ is the inverse image of the set S under the relation $\mathcal{T}_{jk}$. 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.

3.2 Cautious trust

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

$$B_i(I_{ij}\varphi \supset B_j\varphi) \tag{7}$$

$$B_i(B_j\varphi \supset \varphi) \tag{8}$$

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 can not always be sure about the honesty and capability of the agent he would like to trust. For example, according to the past experience, he may know an agent is honest, however, he can not guarantee the agent will keep honest in the future. As for the capability, any agent may make errors even 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 the others if he has good confidence on their honesty and capability. For the agents who sticks to trusting the others only when the two conditions are satisfied, we can call such agents cautious (or strict) ones and their trust is called cautious (or strict) trust. This is an ideal form of trust, so we can define a new class of modal operators T_{ij}^c as

$$T_{ij}^c\varphi =_{def} B_i[(I_{ij}\varphi \supset B_j\varphi) \wedge (B_j\varphi \supset \varphi)]. \tag{9}$$

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

Theorem 3 *For any BIT wffs* φ

1. $\vdash_{\mathrm{BA}} B_iI_{ij}\varphi \wedge T_{ij}^c\varphi \supset B_i\varphi$
2. $\vdash_{\mathrm{BA}} T_{ij}^c\varphi \equiv B_iT_{ij}^c\varphi$

4 Properties of Information Acquisition

4.1 Ideal communication environment

The discussion so far does not pay special attention to the information acquisition operators. However, we still have to clarify some ambiguity for the intuitive meaning of these operators because we do not state the way how the information is acquired. It may mean that i receives a message from j or i has access to the web site where j posts his message. We would like to consider the relationship of $B_iI_{ij}\varphi$ and $I_{ij}\varphi$ in theses two cases.

For the former, when i receives some message from j, he may think that someone pretending j has send the message, so he do not necessarily believe that he has received the message from j. Thus we do not have $I_{ij}\varphi \supset B_iI_{ij}\varphi$. On the other hand, when i receives some message from the address of j and has believed that he indeed received the message from j, it may be really the case that someone else sent it for deception, so $B_iI_{ij}\varphi \supset I_{ij}\varphi$ does not 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 and when he believes j has sent him the message by recognizing the digital signature of j, it is impossible that this is in fact counterfeit by others. Thus we have the following assumption under the ideal environment[3].

C5: $I_{ij}\varphi \equiv B_iI_{ij}\varphi$

For the latter case, the analysis is essentially the same. If the web server is secure enough so that only the owner of the web site can post and update the information (and every agent believes this), then we can also assume C5. The corresponding semantic constraint for C5 is:

(m5) $\mathcal{B}_i \circ \mathcal{I}_{ij} = \mathcal{I}_{ij}$

A basic model satisfying (m5) will be called ideal communication model and the class of such models is denoted by **IC**. The system IC is the result of adding C5 to BA and replace C1 by

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

Under ideal communication environment, if we allow the message-sending interpretation of I_{ij}, then it is possible that $I_{ij}\varphi$ and $I_{kj}\neg\varphi$ hold at the same time. That is, if private communication is allowed, then the agent j may tell one agent the truth but lies to another one. Thus let us consider the situation that an agent can tell others something only by announcing it in public. This is the case when a group of agents subscribe to a mailing list and only communicate with it. In this case, we can add the axiom $I_{ij}\varphi \equiv I_{kj}\varphi$ and require the semantic constraint that for all $1 \leq i, j, k \leq n$, $\mathcal{I}_{ij} = \mathcal{I}_{kj}$ and still have the soundness and completeness results. However, we can even

[3] In fact, even under the imperfect communication environment, a special case of C5, i.e. $I_{ii}\varphi \equiv B_iI_{ii}\varphi$, should still hold intuitively, though it is not included in our basic system for simplicity.

further simplify the language of the BIT logic. For each j, the class of operators $I_{1j}, \ldots, I_{nj}$ can be replaced by an 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. Let C5' denote the axiom $U_j\varphi \equiv B_iU_j\varphi$ and IU denote BU+C5'. Let (m1') and (m5') denote the results of replacing $\mathcal{I}_{ij}$ by $\mathcal{U}_j$ in (m1) and (m5) respectively and let **BU**(resp. **IU**) denote the classes of BUT models satisfying (m1') and (m2)(resp. (m1'), (m2) and (m5')). 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 [11] 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 one and an axiom like C5' holds there. Though the system (called KU there) is different with ours, it is similar with IU here, so we can also define what is 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 *irresponsible liar* if $U_i\varphi \wedge \neg B_i\varphi$ is true. Obviously, an intentional liar is also a irresponsible one. Let $L_i\varphi$ denote i is a irresponsible liar, then we have $\vdash_{\mathrm{IU}} L_i\varphi \supset \neg T_{ii}\varphi$, i.e. a liar can not trust himself (at least in what he is lying).

In the context of IU, an agent i is said to be *honest*[4] if it is not a irresponsible liar, i.e., $U_i\varphi \supset B_i\varphi$ for all φ of BUT logic and *frank* if $B_i\varphi \supset U_i\varphi$. 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

$$B_j\varphi \wedge T_{ij}\varphi \supset B_i\varphi. \tag{10}$$

This means that if i trust j, then i will believe what j believed. 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 conceptually related to the delegation logic proposed in [8] for reasoning about the authorization decision in distributed system.

5 Concluding Remarks

In [1], it is argued that trust is a notion of crucial importance for multi-agent systems. While they 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 the modal logic tools, we characterize the relationship among belief, information acquisition and trust both semantically and axiomatically. In addition to the basic system, some variants and further axioms for trust and information acquisition have been also considered.

A related research direction is about information fusion, where an agent must decide what to accept among possibly inconsistent information from different sources with various degree of reliability. In [3], it is shown how literal information can be merged and suggested that for information of general form, the belief revision approach of Katsuno and Medelzon[7] can be used. In our BIT logic, the notion of trust is only a qualitative concept, so it will be interesting to generalize it to a quantitative notion so that we can merge acquired information from sources with various degrees of reliability.

The second direction is the dynamics of information acquisition. So far, the I_{ij} operators only describe the static facts that some information is acquired. However, we can also consider how the information acquisition action cause the transition of the belief state. Then we can try to develop an update semantics for these operators[14] along the direction of the works reported in [6, 12, 13].

Finally, 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 our logic. Then the trust degree of i on j at time t, denoted by $d^t_{ij} : W \to [0, 1]$ can be defined by

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

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.

[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.

REFERENCES

[1] C. Castelfranchi and R. Falcone, "'Principle of trust for MAS: cognitive anatomy, social importance, and quantification"', in *Proc. of the 3rd International Conference on Multi Agent Systems*, ed., Y. Demazeau, pp. 72–79. IEEE, (1998).

[2] B.F. Chellas, *Modal Logic: An Introduction*, Cambridge University Press, 1980.

[3] L. Cholvy, "'A logiccal approach to multi-souces reasoning"', in *Knowledge Representation and Reasoning under Uncertainty*, eds., M. Masuch and L. Pólos, LNCS 808, pp. 183–196. Springer-Verlag, (1994).

[4] D.C. Dennett, *The Intentional Stance*, MIT Press, Cambridge, MA, 1987.

[5] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi, *Reasoning about Knowledge*, MIT Press, 1996.

[6] J. Gerbrandy and W. Groeneveld, "'Reasoning about information change"', *Journal of Logic, Language, and Information*, **6**, 147–169, (1997).

[7] H. Katsuno and A. Medelzon, "'Propositional knowledge base revision and minimal change"', *Artificial Intelligence*, **52**, 263–294, (1991).

[8] N. Li, B.N. Grosof, and J. Feigenbaum, "'A logic-based knowledge representation for authorization with delegation"', Technical Report R-C21492(96966), IBM, (1999).

[9] J.-J. Ch. Meyer and W. van der Hoek, *Epistemic Logic for AI and Computer Science*, Cambridge University Press, 1995.

[10] Y. Shoham, "'Agent-oriented programming"', *Artificial Intelligence*, **60**(1), 51–92, (1993).

[11] A. Tzouvaras, "'Logic of knowledge and utterance and the liar"', *Journal of Philosophical Logic*, **27**, 85–108, (1998).

[12] B. van Linder, W. van der Hoek, and J.-J.Ch. Meyer, "'Tests as epistemic updates"', in *Proc. of ECAI*, ed., A. Cohen, pp. 331–335. John Wiley & Sons, (1994).

[13] B. van Linder, W. van der Hoek, and J.-J.Ch. Meyer, "'Actions that make you change your mind"', in *Proc. of KI-95*, eds., I. Wachsmuth, C. Rollinger, and W. Brauer, LNAI 981, pp. 185–196. Springer-Verlag, (1995).

[14] F. Veltman, "'Defaults in update semantics"', *Journal of Philosophical Logic*, **25**, 221–261, (1996).

[15] M. Wooldridge and N. Jennings, 'Intelligent agents: theory and practice', *Knowledge Engineering Review*, **10(2)**, 115–152, (1995).