A Logic of Graded Trust and Belief Fusion

Robert Demolombe
ONERA
Centre de Toulouse, DTIM
France
E-mail: Robert.Demolombe@cert.fr

Churn-Jung Liau
Institute of Information Science
Academia Sinica, Taipei
Taiwan
liaucj@iis.sinica.edu.tw

Abstract

In this paper, we propose a logic for integrated reasoning about graded trust and belief fusion in multi-agent systems. Two kinds of graded trust in information sources are analyzed. One is the trust in validity and the other is the trust in completeness. Based on the former, an agent can induce his partial beliefs from what the other agents have told him, whereas by the latter, he can induce his partial beliefs from what the other agents did *not* tell him. Then an agent would have beliefs of different grades when he received information from other agents, so he has to merge these graded beliefs to form his overall belief. The fusion of graded beliefs may be done according to two different strategies. In the first one, if inconsistency occurs at some level, then all beliefs at the lower levels are discarded simultaneously, so it is called level cutting strategy. For the second one, only the level at which the inconsistency occurs is skipped, so it is called level skipping strategy. It is shown that these notions can all be formulated in a modal logic framework. The syntax, semantics and axiomatic systems for such logical formulation are presented in this paper.

1 Introduction

Recently, the research of intelligent agents have received much attention due to the rapid development of internet applications and electronic commerce. An intelligent agent can search through the web and find information matching the user's need or buy and sell goods on behalf of the users[13]. In most cases, agents do not live alone. They are situated in a multi-agent societies. 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. It has been shown that the notions of trust play a very important role in the interaction between agents, whether they are competitive or cooperative[3].

In [14], 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 [7], 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, such as epistemic logic, doxastic logic[9], and deontic logic[2]. Some logics derived from the philosophical analysis have been applied to the modeling of AI and distributed systems[12, 8]. 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[4]. In [6, 11], the notion of trust is formulated as a kind of modal operators and its relationship with agents' belief and information transmission action is characterized by logical axioms. These systems formulate a dichotomous concept of trust, thus when agent i received contradictory information from agents j and k, he can not trust both of them. In [6], the formulation is further extended to graded trust, where we can speak of an agent i trusting j to some level. However, in the formulation, if agent i trusts agents j and k to different levels respectively, then his belief will crash when he received contradictory information from them. In the practice, however, an agent may induce graded belief from the graded trust and the information he has received, so if he received contradictory information from two (or more) agents whom he only partially trust, it should be possible for the agent to merge the induced graded belief and form his overall belief.

In [10], a logic of belief fusion based on multi-agent epistemic logic[8] and multi-sources reasoning[5] has been proposed. While the logic merges the beliefs of different agents according to a total ordering on the reliability of these agents, it can also be applied to fusion of beliefs at different levels for a single agent. There are two strategies for the fusion of beliefs in [10]. In the first one, if inconsistency occurs at some level, then all beliefs at the lower levels are discarded simultaneously, so it is called level cutting strategy. For the second one, only the level at which the inconsistency occurs is skipped, so it is called level skipping strategy.

Thus, following these approaches, we would like to propose a doxastic logic integrating graded trust and belief fusion, and then discuss how one agent's belief is influenced by his graded trust toward other agents and the information he acquires. More specifically, in traditional doxastic logic, $K_i\varphi$ means that agent i believes φ, so we will add to the logic additional modal operators B_i^α, TV_{ij}^α, TC_{ij}^α and I_{ij}. The operator B_i^α is for the graded belief of agent i at level α and the intended meaning of $TV_{ij}^\alpha\varphi$(resp. TC_{ij}^α) is that agent i trusts the validity(resp. completeness) of j on φ to some extent α, whereas $I_{ij}\varphi$ means agent i acquires information φ from j. The logic of graded trust then characterizes the relationship between graded trust, information acquisition action, and graded belief. Then belief fusion operators based on two different strategies are introduced to merge the graded beliefs for each agent and the resultant merged belief will be related with the agents' ordinary belief modeled by the K_i operators.

In the rest of the paper, we will first give a general logic for graded trust. The syntax, semantics, and a basic axiomatic system of the logic is presented in section 2. The logic of belief fusion based on level cutting and skipping strategies are presented in sections 3 and 4 respectively. In section 5, we consider some practical problems in the application of the logic. Finally, we conclude the paper with a brief summary of the results.

2 A Logic of Graded Trust

Let Λ be a finite set of symbols, called grades or levels, and assume there are n agents and a set Φ_0 of countably many atomic propositions, then the set of well-formed formulas(wff) for the logic of graded trust(LGT) under Λ is the least set containing Φ_0 and closed under the following formation rules:

- if φ is a wff, so are $\neg\varphi$, $K_i\varphi$, $B_i^\alpha\varphi$, $I_{ij}\varphi$, $TV_{ij}^\alpha\varphi$ and $TC_{ij}^\alpha\varphi$ for all $1 \leq i \neq j \leq n$ and $\alpha \in \Lambda$,
- 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 doxastic operator K_i is to model the belief of agents, so it is a kind of KD45 normal modal operators[4]. In [6], $K_i\varphi$ is read as "agent i strongly believes φ". Since in this paper, an agent's belief is related with the fusion of information he has received from other agents, we can say that agent i deliberatively believes φ if $K_i\varphi$ holds. However, for simplicity, we will in general only read $K_i\varphi$ as "agent i believes φ". On the other hand, as in [6], B_i^α is to model the graded beliefs of agents, so $B_i^\alpha\varphi$ means that agent i believes φ at the level α. Following [6], we require that it is a KD normal modal operator. An agent's graded beliefs are closely related to his graded trust in other agents and the information he has received (or not received) from them. There are at least six notions of trust mentioned in [6]. These are trusts on the agent's validity, completeness, sincerity, credibility, cooperativity, and vigilance. Here we will consider the first two, that is, validity and completeness. We choose these two notions because they directly connect the fact sent (or not) by an agent to the receiving agent's belief on the same fact. This makes it possible to model the relationship between an agent's acquired information and its belief. An agent is valid on a proposition if his affirmation of the proposition implies it is true and complete on it if his not asserting it means it is false. These notions are respectively modeled by TV_{ij}^{α}, TC_{ij}^{α} and I_{ij}. Thus $TV_{ij}^{\alpha}\varphi$(resp. TC_{ij}^{α}) means that agent i trusts the validity(resp. completeness) of j in φ to some extent α, whereas $I_{ij}\varphi$ means that j has informed i of φ or i has acquired information φ from j.

The possible-worlds semantics provides a general framework for the modeling of knowledge and belief[8]. 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 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.

According to the informal discussion above, the formal semantics for K_i, B_i^{α} and I_{ij} is the Kripke semantics for normal modal operators, whereas that for TV_{ij}^{α} and TC_{ij}^{α} is the so-called minimal (or neighborhood) semantics[4]. Formally, a LGT model is a tuple

$$(W, \pi, (\mathcal{K}_i)_{1\leq i\leq n}, (\mathcal{B}_i^{\alpha})_{\alpha\in\Lambda,1\leq i\leq n}, (\mathcal{I}_{ij})_{1\leq i\neq j\leq n}, (\mathcal{TV}_{ij}^{\alpha})_{\alpha\in\Lambda,1\leq i\neq j\leq n}, (\mathcal{TC}_{ij}^{\alpha})_{\alpha\in\Lambda,1\leq i\neq j\leq n})$$

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,
- $\mathcal{K}_i \subseteq W \times W$ is a serial, transitive, and Euclidean binary relation[1] on W
- $\mathcal{B}_i^{\alpha} \subseteq W \times W$ and $\mathcal{I}_{ij} \subseteq W \times W$ are both serial binary relations on W,
- $\mathcal{TV}_{ij}^{\alpha} \subseteq W \times 2^W$ and $\mathcal{TC}_{ij}^{\alpha} \subseteq W \times 2^W$ are binary relations between W and the power set of W.

For simplicity, we will write an LGT model as $(W, \pi, \mathcal{K}_i, \mathcal{B}_i^{\alpha}, \mathcal{I}_{ij}, \mathcal{TV}_{ij}^{\alpha}, \mathcal{TC}_{ij}^{\alpha})$ with the understanding that $1 \leq i \neq j \leq n$ and $\alpha \in \Lambda$.

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 $w \in W$, $\mathcal{K}_i(w)$, $\mathcal{B}_i^{\alpha}(w)$ and $\mathcal{I}_{ij}(w)$ will be subsets of W, whereas $\mathcal{TV}_{ij}^{\alpha}(w)$ and $\mathcal{TC}_{ij}^{\alpha}$ are subsets of 2^W. Informally, $\mathcal{K}_i(w)$(resp. $\mathcal{B}_i^{\alpha}(w)$) is the set of worlds that agent i considers possible under w according to his belief (resp. at level α), whereas $\mathcal{I}_{ij}(w)$ is that agent i considers possible according to the

---

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

information acquired from j. On the other hand, since each subset of W is the semantic counterpart of a proposition, for any $S \subseteq W$, $S \in \mathcal{TV}_{ij}^{\alpha}(w)$ (resp. $S \in \mathcal{TC}_{ij}^{\alpha}(w)$) means that agent i trust j's validity (resp. completeness) to the extent α on the truth of the proposition corresponding to S. The informal intuition is reflected in our formal definition of satisfaction relation. Let $M = (W, \pi, \mathcal{K}_i, \mathcal{B}_i^{\alpha}, \mathcal{I}_{ij}, \mathcal{TV}_{ij}^{\alpha}, \mathcal{TC}_{ij}^{\alpha})$ be an LGT 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 $w \in \pi(p)$ if $p \in \Phi_0$,
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. $w \models K_i\varphi$ iff for all $u \in \mathcal{K}_i(w)$, $u \models \varphi$,
5. $w \models B_i^{\alpha}\varphi$ iff for all $u \in \mathcal{B}_i^{\alpha}(w)$, $u \models \varphi$,
6. $w \models I_{ij}\varphi$ iff for all $u \in \mathcal{I}_{ij}(w)$, $u \models \varphi$,
7. $w \models TV_{ij}^{\alpha}\varphi$ iff $|\varphi| \in \mathcal{TV}_{ij}^{\alpha}(w)$, where $|\varphi| = \{u \in W : u \models \varphi\}$ is called the truth set of φ,
8. $w \models TC_{ij}^{\alpha}\varphi$ iff $|\varphi| \in \mathcal{TC}_{ij}^{\alpha}(w)$.

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 LGT 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 an LGT model so that the relations $\mathcal{B}_i^{\alpha}$,$\mathcal{I}_{ij}$, $\mathcal{TV}_{ij}^{\alpha}$ and $\mathcal{TC}_{ij}^{\alpha}$ 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 LGT models. Let $M = (W, \pi, \mathcal{K}_i, \mathcal{B}_i^{\alpha}, \mathcal{I}_{ij}, \mathcal{TV}_{ij}^{\alpha}, \mathcal{TC}_{ij}^{\alpha})$ be an LGT model, then M is called *ideal* if it satisfies the following constraints for all $\alpha \in \Lambda, 1 \leq i \neq j \leq n$ and $w \in W$,

(m1.1) for all $S \in \mathcal{TV}_{ij}^{\alpha}(w)$, if $\mathcal{K}_i(w) \subseteq \{u \mid \mathcal{I}_{ij}(u) \subseteq S\}$,[2] then $\mathcal{B}_i^{\alpha}(w) \subseteq S$,

(m1.2) for all $S \in \mathcal{TC}_{ij}^{\alpha}(w)$, if $\mathcal{K}_i(w) \subseteq \{u \mid \mathcal{I}_{ij}(u) \cap \overline{S} \neq \emptyset\}$, then $\mathcal{B}_i^{\alpha}(w) \subseteq \overline{S}$,

(m2.1) for all $w, u, v \in W$, $\mathcal{K}_i(w, u)$ and $\mathcal{I}_{ij}(u, v)$ implies $\mathcal{I}_{ij}(w, v)$,

(m2.1) for all $w, u, v \in W$, $\mathcal{K}_i(w, u)$ and $\mathcal{I}_{ij}(w, v)$ implies $\mathcal{I}_{ij}(u, v)$.

The class of ideal LGT models is denoted by **IG**. The constraints (m2.1) and (m2.2) essentially require that the information transmission action between agents occurs in an ideal communication environment, so once agent i acquires information from j, he believes the interaction indeed has occurred and vice verse. On the other hand, (m1.1) and (m1.2) make a connection among the three classes of modal operators. (m1.1) means that if an agent i believes he has acquired the information φ from j and he trusts the validity of j to the extent α on the truth of φ, then he should believe φ at the level α. (m1.2) means that if an agent i believes j did not informed him of φ and he trusts the completeness of j to the extent α on the truth of φ, then he should believe $\neg\varphi$ at the level α. These constraints are represented by natural axioms in our axiomatic system for ideal LGT logic. The axiomatic system, called IG, is presented in Fig. 1.

---

[2] This is equivalent to $\mathcal{K}_i \circ \mathcal{I}_{ij} \subseteq S$, where $\circ$ denotes the relational composition.

1. Axioms:

   P: all tautologies of the propositional calculus

   K1: $[K_i\varphi \wedge K_i(\varphi \supset \psi)] \supset K_i\psi$

   K2: $\neg K_i\bot$

   K3: $K_i\varphi \supset K_iK_i\varphi$

   K4: $\neg K_i\varphi \supset K_i\neg K_i\varphi$

   B1: $[B_i^\alpha\varphi \wedge B_i^\alpha(\varphi \supset \psi)] \supset B_i^\alpha\psi$

   B2: $\neg B_i^\alpha\bot$

   I1: $[I_{ij}\varphi \wedge I_{ij}(\varphi \supset \psi)] \supset I_{ij}\psi$

   I2: $\neg I_{ij}\bot$

   C1.1: $TV_{ij}^\alpha\varphi \supset (K_iI_{ij}\varphi \supset B_i^\alpha\varphi)$

   C1.2: $TC_{ij}^\alpha\varphi \supset (K_i\neg I_{ij}\varphi \supset B_i^\alpha\neg\varphi)$

   C2.1: $I_{ij}\varphi \supset K_iI_{ij}\varphi$

   C2.2: $\neg I_{ij}\varphi \supset K_i\neg I_{ij}\varphi$

2. Rules of Inference:

   R1(Modus ponens, MP): from $\vdash \varphi$ and $\vdash \varphi \supset \psi$ infer $\vdash \psi$

   R2(Necessitation, Nec): from $\vdash \varphi$ infer $\vdash K_i\varphi$, $\vdash B_i^\alpha\varphi$, and $\vdash I_{ij}\varphi$

   R3: from $\vdash \varphi \equiv \psi$ infer $\vdash TV_{ij}^\alpha\varphi \equiv TV_{ij}^\alpha\psi$ and $\vdash TC_{ij}^\alpha\varphi \equiv TC_{ij}^\alpha\psi$

Figure 1: The axiomatic system IG for ideal LGT

The axioms K1-K4 correspond to the KD45 system for doxastic operators K_i, K1 means that the agents are perfect logical reasoners, so their belief are closed under logical consequence. K2-K4, corresponding to the serial, transitive and Euclidean properties of the $\mathcal{K}_i$ relation, stipulate respectively the consistency, positive introspection, and negative introspection of the agent's belief. Analogously, the axioms B1-B2 and I1-I2 correspond to the KD systems for doxastic operators B_i^α and the information acquisition operators I_{ij}. Here, we assume that the information acquisition operators describe not only the explicit information an agent acquires directly but also all consequences that are implicitly implied by it, so if an agent acquires the information φ, he also obtains all logical consequences of φ at the same time. This is just what I1 asserts. Though there may be differences between the epistemic status of the externally acquired explicit information and that of the internally reasoned consequences, we do not have to model both of them because it is the latter notion that is really related to the agent's belief when we consider ideally perfect reasoning agents. 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 graded trust can help to select what to believe when such situation occurs. Finally, the connection axioms C1.1 to C2.2 correspond to the same numbered constraints on the ideal LGT models. C1.1 and C1.2 specify the characteristic properties of trust in validity and completeness respectively, whereas C2.1 and C2.2 describe that the agent is aware of the receipt or not of the information. By axioms C2.1 and C2.2, axioms C1.1 and C1.2 can in fact be simplified respectively to

$$TV_{ij}^\alpha \varphi \supset (I_{ij}\varphi \supset B_i^\alpha \varphi)$$

and

$$TC_{ij}^\alpha \varphi \supset (\neg I_{ij}\varphi \supset B_i^\alpha \neg\varphi).$$

However, under imperfect communication environment, if i received some information from j but j can not prove himself to i (so i is not sure he has received information from j, i.e. $I_{ij}\varphi$ but $\neg K_i I_{ij}\varphi$), then the trust of i in j is not sufficient for i to form his belief.

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

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

The system IG is not in much difference with that for dichotomous trust in [11] except there is now a class of doxastic operators instead of a single B_i for each agent i. What makes the difference will be the fusion of an agent's beliefs in different levels into a total belief. In fact, in the systems of graded trust in [6], a partial ordering $\geq$ is imposed on the set of levels and the following two axioms are considered

$$B_i^\alpha \varphi \supset B_i^\beta \varphi \quad \text{if} \quad \alpha \geq \beta$$

$$K_i \varphi \supset B_i^\alpha \varphi \quad \text{for all} \quad \alpha \in \Lambda$$

where $K_i\varphi$ denotes that agent i strongly believes φ. This is based on the interpretation of $B_i^\alpha\varphi$ as "agent i believes φ at or above the level α". By this interpretation, if agent i received information φ from j and $\neg\varphi$ from k, then $TV_{ij}^\alpha\varphi$ and $TV_{ik}^\beta\neg\varphi$ can not hold simultaneously if α and β has a common lower bound in Λ since otherwise, from C1.1 and the first axiom above, it is contradictory with the consistency of belief B2. Therefore, to utilize the ordering on the levels, we would like to interpret $B_i^\alpha \varphi$ as agent i believes φ *exactly* at the level α. By adopting this interpretation, then it is possible that for an agent i, $B_i^\alpha \varphi$ and $B_i^\beta \neg\varphi$ hold simultaneously without contradiction. Thus it becomes possible to merge the agent's beliefs at different levels.

There are several possibilities to achieve belief fusion. The most straightforward way is just to put them together. This can be done for any subsets of levels, so we will introduce a new class of doxastic operators B_i^G for each agent i and $G \subseteq \Lambda$. Intuitively, $B_i^G \varphi$ means that φ can be deduced by pooling together the beliefs of agent i at all levels belonging to G, so it is also called *direct merging.* However, since an agent may receive contradictory information from different sources, just putting them together would result in the crash of the agent's merged belief. Thus more sophisticated belief fusion approaches based on the partial ordering of levels are expected. In [10], two conservative belief fusion strategies for multi-agent systems are proposed. In the first one, if inconsistency occurs at some level, then all beliefs at the lower levels are discarded simultaneously, so it is called level cutting strategy. For the second one, only the level at which the inconsistency occurs is skipped, so it is called level skipping strategy. While the belief fusion approach considered in [11] is based on total orders on agents, we have to merge an agent's belief according to a partial ordering on the set of levels. This can be achieved by considering all total orders compatible with the given partial ordering. We will describe the approach in the next two sections.

3 Fusion by Level Cutting

Let $\succ$ be a strict partial ordering on Λ, i.e., a binary relation on Λ satisfying irreflexivity and transitivity[3], then a total order compatible with $(\Lambda, \succ)$ is any $(\Delta, >)$ such that $\Delta \subseteq \Lambda$, $>$ is a strict total ordering on Δ, and for any $\alpha, \beta \in \Delta$, $\alpha \succ \beta$ implies $\alpha > \beta$. The set of total orders compatible with $(\Lambda, \succ)$ is denoted by $\mathcal{O}_\succ(\Lambda)$. If $\Delta = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$, then we usually write the total order $(\Delta, >)$ as $\alpha_1 > \alpha_2 > \cdots > \alpha_k$. Let us denote the elements of $\mathcal{O}_\succ(\Lambda)$ by O and if $O = \alpha_1 > \alpha_2 > \cdots > \alpha_k$, then $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ is called the domain of O and is denoted by $\delta(O)$. Furthermore, in this case, $O > \alpha_{k+1}$ denotes the total order $\alpha_1 > \alpha_2 > \cdots > \alpha_k > \alpha_{k+1}$. Also the cardinality of $\delta(O)$(i.e. the number k) is called the length of O and denoted by $len(O)$. The symbol $\mathcal{O}_\succ^k(\Lambda)$ denotes the subset $\{O \in \mathcal{O}_\succ(\Lambda) \mid len(O) = k\}$. Sometimes $\succ$ and Λ are omitted in these notations when they are clear from the context.

To model the fusion of agent's belief, we add the new modal operators $B_i^G \varphi$ and $B_i^O \varphi$ for each agent i, $\emptyset \neq G \subseteq \Lambda$ and $O \in \mathcal{O}$ and the following formation rule to LGT

- if φ is wff, so are $B_i^G \varphi$ and $B_i^O \varphi$ for each $1 \leq i \leq n$, $\emptyset \neq G \subseteq \Lambda$ and $O \in \mathcal{O}$.

Intuitively, $B_i^O \varphi$ means that φ is deducible from merging the agent i's graded belief according to the total ordering O, whereas $B_i^G \varphi$ is the result of directly putting the agent's beliefs together which has been mentioned above. In the following, we will somewhat abuse the notations and identify B_i^α with B_i^G and B_i^O when $G = \{\alpha\}$ and O is the unique total ordering on $\{\alpha\}$. The fusion of belief under a total ordering in this section is according to the level cutting strategy, so in the semantic side, we would define the binary relations $\mathcal{B}_i^G \subseteq W \times W$ and $\mathcal{B}_i^O \subseteq W \times W$ by

$$\mathcal{B}_i^G = \bigcap_{\alpha \in G} \mathcal{B}_i^\alpha \tag{1}$$

and

$$\mathcal{B}_i^{O>\alpha}(w) = \begin{cases} \mathcal{B}_i^O(w) & \text{if } \bigcap_{\beta \in \delta(O>\alpha)} \mathcal{B}_i^\beta(w) = \emptyset, \\ \mathcal{B}_i^O(w) \cap \mathcal{B}_i^\alpha(w) & \text{otherwise,} \end{cases} \tag{2}$$

---

[3] $\neg(x \succ x)$ and $x \succ y \wedge y \succ z \Rightarrow x \succ z$

for any $w \in W$, and the satisfaction relation is extended to the wffs $B_i^G\varphi$ and $B_i^O\varphi$ by

1. $w \models B_i^G\varphi$ iff for all $u \in \mathcal{B}_i^G(w)$, $u \models \varphi$, and
2. $w \models B_i^O\varphi$ iff for all $u \in \mathcal{B}_i^O(w)$, $u \models \varphi$.

According to semantics of B_i^G operators, $\mathcal{B}_i^G(w)$ is the set of worlds agent i considers possible by putting together all of his belief at levels belonging to G, so it should be the intersection of the $\mathcal{B}_i^\alpha(w)$'s for all $\alpha \in G$. If the intersection is empty, it means that there is a contradiction by putting together these belief sets directly. Thus in the definition of $\mathcal{B}_i^{O>\alpha}(w)$, if there is no contradiction by directly putting together all belief sets at levels belonging to the domain of $O > \alpha$, then we can just do it, otherwise, we have to recursively reduce it to $\mathcal{B}_i^O(w)$. If $O = \alpha_1 > \cdots > \alpha_k > \alpha_{k+1} > \cdots$, and $\bigcap_{1\leq j\leq k} \mathcal{B}_i^{\alpha_j}(w) \neq \emptyset$ but $\bigcap_{1\leq j\leq k+1} \mathcal{B}_i^{\alpha_j}(w) = \emptyset$, then obviously, $\mathcal{B}_i^O(w) = \bigcap_{1\leq j\leq k} \mathcal{B}_i^{\alpha_j}(w)$, so the level below α_k is cut out. This is why it is named level cutting strategy. Since the belief sets at different levels are merged according to a total ordering whereas we have only a partial ordering $\succ$ on Λ, we would consider all total orders of length $|\Lambda|$ that are compatible with $\succ$. Thus an agent will believe φ (with deliberations) if it is a consequence of all these fusion results. This is semantically achieved by the following constraint.

Let $|\Lambda| = m$, then an FC model (meaning "fusion by cutting strategy") is an ideal LGT model satisfying

(m3) $\mathcal{K}_i \subseteq \bigcup_{O\in\mathcal{O}^m} \mathcal{B}_i^O$.

The class of FC models is denoted by **FC**. The axiomatic system FC is the extension of IG with the axioms in Fig. 2 and the Nec rule for B_i^G. Then we have

Theorem 2 *The system* FC *is sound and complete with respect to* **FC**.

The axioms G1 and the Nec rule for B_i^G assert that it is still a normal modal operator. However, note that except when $|G| = 1$, the consistency property $\neg B_i^G\bot$ does not hold any more since the direct merging of belief may result in contradiction. G2 says that the more levels of belief sets are merged, the more consequences can be derived. These axioms correspond exactly to those for distributed knowledge in multi-agent epistemic logic[8]. There is only a superficial difference that the distributed knowledge is to merge knowledge of different agents whereas here we do it for a single agent's belief at different levels. The axioms O1 and O2 reflect exactly two conditions on the definition of $\mathcal{B}_i^{O>\alpha}(w)$. O1 is the case where direct merging does not cause contradiction, i.e., all levels of belief sets in the domain of $O > \alpha$ are consistent, then the fusion result is equivalent to the direct merging one. However, if contradiction indeed occurs due to direct merging, then the lowest level of belief must be discarded first and then the fusion result is recursively reduced to that for the remaining levels. This is just what O2 asserts. Finally, the axiom C3 ties the fusion results of a single agent with his overall belief. This means that if a consequence can be derived from the fusion results under all possible total orders compatible with $\succ$, then the agent would believe it. Note that the converse of C3 does not hold because we would not like to exclude the possibility that an agent can acquire his belief from outside of the information he has received from other agents(e.g. by observation made by himself).

4 Fusion by Level Skipping

Admittedly, the level cutting strategy for information fusion is very conservative since it discards all information below a level where the inconsistency occurs. A less conservative one is to just skip the level causing inconsistency and continue absorbing the belief below it into the fusion result if they are after all consistent by doing so. To model this kind of fusion strategy, we have to add to our language the modal operators B_i^Ω for any nonempty $\Omega \subseteq \mathcal{O}_\succ(\Lambda)$ and the rule:

G1: $(B_i^G\varphi \wedge B_i^G(\varphi \supset \psi)) \supset B_i^G\psi$

G2: $B_i^{G_1}\varphi \supset B_i^{G_2}\varphi$ if $G_1 \subset G_2$

O1: $\neg(B_i^{\delta(O>\alpha)}\bot) \supset (B_i^{O>\alpha}\varphi \equiv B_i^{\delta(O>\alpha)}\varphi)$

O2: $(B_i^{\delta(O>\alpha)}\bot) \supset (B_i^{O>\alpha}\varphi \equiv B_i^{O}\varphi)$

C3: $(\bigwedge_{O\in\mathcal{O}^m} B_i^O\varphi) \supset K_i\varphi$

Figure 2: The axioms to be added for FC

- if φ is a wff, so is $B_i^\Omega\varphi$ for any nonempty $\Omega \subseteq \mathcal{O}$.

Note that each $\alpha \in \Lambda$ can also be seen as the unique total ordering on the singleton $\{\alpha\}$, so any subset of Λ is in fact also a subset of $\mathcal{O}$. This means that each B_i^G where $G \subseteq \Lambda$ can be seen as a special case of the more general operator B_i^Ω. Also, by abusing the notation, B_i^O can be identified with $B_i^{\{O\}}$.

For the semantics, an FC model can still serve as an FS (meaning fusion by skipping strategy) model, however, the inductive definition of $\mathcal{B}_i^O$ in (2) must be modified as follows:

$$\mathcal{B}_i^{O>\alpha}(w) = \begin{cases} \mathcal{B}_i^O(w) & \text{if } \mathcal{B}_i^O(w) \cap \mathcal{B}_i^\alpha(w) = \emptyset, \\ \mathcal{B}_i^O(w) \cap \mathcal{B}_i^\alpha(w) & \text{otherwise,} \end{cases} \tag{3}$$

for any $w \in W$. Then a definition for $\mathcal{B}_i^\Omega \subseteq W \times W$ is

$$\mathcal{B}_i^\Omega = \bigcap_{O\in\Omega} \mathcal{B}_i^O \tag{4}$$

In the definition above, when a new level α is added to the tail of a total ordering O, the agent will decide whether accept the belief at that level by considering its consistency with those already merged according to the ordering O instead of the consistency of directly merging these levels of belief sets. Therefore, if there are some levels in O which have been discarded due to inconsistency with higher levels of belief, then these discarded levels do not have any effects on $\mathcal{B}_i^O(w)$, so they does not influence the test for the acceptance of the level α. In effect, these discarded levels have been skipped. However, testing the emptiness of $\mathcal{B}_i^O(w) \cap \mathcal{B}_i^\alpha(w)$ is equivalent to check the consistency of directly merging the fusion result according to the ordering O with the belief at level α. This is the reason why we need the operators of the form B_i^Ω which can directly merge the fusion results according to different total orders in Ω.

The satisfaction relation is extended to $B_i^\Omega\varphi$ as usual. An FS model must still satisfy (m3) though the $\mathcal{B}_i^O$ appearing there is now according to the new definition (3). The class of FS models is denoted by **FS** and the axiomatic system FS consists of IG (the Nec rule is rewritten as "from $\vdash \varphi$ infer $\vdash K_i\varphi$, $\vdash B_i^\Omega\varphi$, and $\vdash I_{ij}\varphi$"), axiom C3, and those in Fig. 3.

The axioms Ω1 and Ω2 are analogous to G1 and G2 except that the levels are replaced by total orders. Axioms O1' and O2' are for the fusion by skipping strategy. The Ω in these two axioms may be empty or nonempty subsets of $\mathcal{O}$. When Ω is empty, they correspond exactly to the two conditions on (3). However, for the logical completeness of the system, we have to add the general term Ω in them[10].

Theorem 3 *The system* FS *is sound and complete with respect to* **FS**.

Ω1: $(B_i^{\Omega}\varphi \wedge B_i^{\Omega}(\varphi \supset \psi)) \supset B_i^{\Omega}\psi$

Ω2: $B_i^{\Omega_1}\varphi \supset B_i^{\Omega_2}\varphi$ if $\Omega_1 \subset \Omega_2$

O1': $\neg(B_i^{\{O,\alpha\}}\bot) \supset (B_i^{\Omega\cup\{O>\alpha\}}\varphi \equiv B_i^{\Omega\cup\{O,\alpha\}}\varphi)$

O2': $(B_i^{\{O,\alpha\}}\bot) \supset (B_i^{\Omega\cup\{O>\alpha\}}\varphi \equiv B_i^{\Omega\cup\{O\}}\varphi)$

Figure 3: The additional axioms for FS

5 Some Practical Consideration

A practical problem to be considered when applying our logic to the derivation of the agents' belief is regarding axioms O1 and O1'. Though these two axioms are heavily used in the fusion of agent's graded belief, their antecedents (i.e., wffs of the form $\neg B_i^G\bot$ or $\neg B_i^{\Omega}\bot$) are hardly derivable from the mere description of a situation.

Example 1 Assume a set of grades $\Lambda = \{1, 2, 3\}$ with the partial ordering $1 \succ 2 \succ 3$ is given and the situation is described by the following set of wffs,

$$\Sigma = \{TV_{ij}^1 p, TV_{ik}^2 q, TV_{ik}^3 \neg p, I_{ij}p, I_{ik}(\neg p \wedge q)\}$$

then by axiom C1.1, the following graded beliefs for agent i can be derived

$$\{B_i^1 p, B_i^2 q, B_i^3 \neg p\}.$$

Obviously, we would like to derive $K_i(p\wedge q)$ since the beliefs in the first two levels are consistent. However, by applying O2 (and other axioms like G1, G2, etc.), though we can prove $\vdash_{\mathrm{FC}} B_i^{1\succ 2\succ 3}(p\wedge q) \equiv B_i^{1\succ 2}(p\wedge q)$ and $\vdash_{\mathrm{FC}} B_i^{\{1,2\}}(p\wedge q)$, we can not prove $\vdash_{\mathrm{FC}} B_i^{1\succ 2}(p\wedge q) \equiv B_i^{\{1,2\}}(p\wedge q)$ because $\neg B_i^{\{1,2\}}\bot$ is not derivable from the premises. This is due to the fact that while we have the positive knowledge about the graded beliefs of i (i.e. what i believes), we do not have the negative one (i.e. what i does not believe). ■

To solve the representation problem of negative knowledge[4], many non-monotonic reasoning mechanisms have been proposed in the AI literature[1]. Here we will adopt a similar but less complicated solution. The basic idea is to add some default assumptions to the set of premises. However, the main difficulty in non-monotonic reasoning is that the default assumptions to be considered are in general infinite and there are hardly constructive approaches to add them (witnessed by the fact that many non-monotonic reasoning mechanisms resort to some kind of fixed point definition). Fortunately, what we are concerning here are only the wffs of the form $\neg B_i^G\bot$ for some agent i and nonempty $G \subseteq \Lambda$. Thus we will restrict the default assumptions to wffs of this form and called them *belief-consistency* literals, or in short, *bc-literals*. The number of bc-literals is finite and actually it is equal to $n \cdot (2^m - 1)$ if $|\Lambda| = m$. Let Σ be a set of wffs in FC(resp. FS) logic, then Σ' is a *bc-extension* of Σ if $\Sigma \subseteq \Sigma'$ and $\Sigma' - \Sigma$ is a maximal subset of bc-literals such that $\Sigma' \nvdash_{\mathrm{FC}} \bot$ (resp. $\Sigma' \nvdash_{\mathrm{FS}} \bot$). Let L= FC or FS and $\Sigma \cup \{\varphi\}$ be a set of wffs for L, then we write $\Sigma \mid\!\sim_{\mathrm{L}}^{\forall} \varphi$ if for every bc-extension Σ' of Σ, $\Sigma' \vdash_{\mathrm{L}} \varphi$ and $\Sigma \mid\!\sim_{\mathrm{L}}^{\exists} \varphi$ if there exists a bc-extension Σ' of Σ such that $\Sigma' \vdash_{\mathrm{L}} \varphi$.

Example 2 (continued) There is exactly one bc-extension to Σ in example 1, it is

$$\Sigma \cup \{\neg B_i^{\{1,2\}}\bot, \neg B_i^{\{1,3\}}\bot, \neg B_j^{\{1,2,3\}}\bot, \neg B_k^{\{1,2,3\}}\bot\},$$

---

[4] Also called frame problem in the AI literature.

so the desired result $K_i(p \wedge q)$ can be finally derived from the extension. Note that we omit some subsumed bc-literals in the bc-extension. For example, from $\neg B_j^{\{1,2,3\}}\bot$, we can derive $\neg B_j^{\{1,2\}}\bot$ by G2, so the explicit incorporation of the latter is not necessary. Thus we have $\Sigma \mid\!\sim^{\forall}_{\mathrm{FC}} K_i(p \wedge q)$ and $\Sigma \mid\!\sim^{\exists}_{\mathrm{FC}} K_i(p \wedge q)$. In this example, we can derive the same results for FS system since the inconsistency occurs at the lowest level. ■

In the example above, the set of premises has only one bc-extension. However, in the general case, there may be a dilemma on the choice between the $\mid\!\sim^{\forall}$ or $\mid\!\sim^{\exists}$ derivability relations. The former is computationally more difficult than the latter since it requires the computation of all bc-extensions of the premises. However, the results derived by the $\mid\!\sim^{\exists}$ relation may be incompatible with each other. The good news is that in many real situation, we do not have to face the dilemma. Let us define a simple situation description(SSD) as a set of wffs of the forms $TV_{ij}^{\alpha}\varphi$, $I_{ij}\varphi$, or $B_i^{\alpha}\varphi$ where φ is a classical propositional wff, then it can be easily shown that

Proposition 1 *If Σ is an SSD such that $\Sigma \not\vdash_{\mathrm{IG}} \bot$, then Σ has exactly one bc-extension.*

In the case of SSD's, the relation $\mid\!\sim^{\forall}$ and $\mid\!\sim^{\exists}$ coincide, so it will be denoted by $\mid\!\sim$ only.

The preceding example is based on a total ordered grades set. Let us now consider a situation where the ordering on the grades is properly partial.

Example 3 Let $\Lambda = \{1, 2, 3, 4\}$ be endowed with the partial order $\{1 \succ 2, 1 \succ 3, 1 \succ 4, 2 \succ 4, 3 \succ 4\}$ and the set of premises be an SSD

$$\Sigma = \{TV_{ij}^{1}(p \vee q), TV_{ij}^{4}r, TV_{ik}^{2}\neg p, TV_{ik}^{3}\neg q, I_{ij}((p \vee q) \wedge r), I_{ik}(\neg p \wedge \neg q)\},$$

then the following graded beliefs are derivable by axiom C1.1.

$$B_i^1(p \vee q), B_i^2\neg p, B_i^3\neg q, B_i^4 r$$

Thus the unique bc-extension is to add the following bc-literals into Σ

$$\neg B_i^{\{1,2,4\}}\bot, \neg B_i^{\{1,3,4\}}\bot, \neg B_i^{\{2,3,4\}}\bot, \neg B_j^{\{1,2,3,4\}}\bot, \neg B_k^{\{1,2,3,4\}}\bot$$

From this extension, we can derive

$$B_i^{1>2>3>4}(p \oplus q) \wedge B_i^{1>3>2>4}(p \oplus q)$$

in both FC and FS, where $\oplus$ denotes the 'exclusive-or' operation and

$$B_i^{1>2>3>4}r \wedge B_i^{1>3>2>4}r$$

in FS, so by C3 and the fact that $\mathcal{O}_{\succ}^{4}(\Lambda) = \{1 > 2 > 3 > 4, 1 > 3 > 2 > 4\}$, we can finally prove

$$\Sigma \mid\!\sim_{\mathrm{FC}} K_i(p \oplus q) \text{ and } \Sigma \mid\!\sim_{\mathrm{FS}} K_i((p \oplus q) \wedge r).$$

Note that the example also shows the difference between level cutting and skipping strategies. The wff r is not accepted by agent i in the FC system because it is believed at level 4 which is lower than the one at which the inconsistency occurs, i.e. levels 2 or 3, whereas for the FS system, no matter which of level 2 or 3 is skipped, the belief at level 4 is still acceptable since it is consistent with the fusion results of beliefs at levels 1 and 2 or 1 and 3 (as assured by the bc-literals $\neg B_i^{\{1,2,4\}}\bot$ and $\neg B_i^{\{1,3,4\}}\bot$).

Note that from the example (and the preceding ones), we can also see why the operators I_{ij} must be the normal ones. If we can not infer the implicit information $I_{ik}\neg p$ and $I_{ik}\neg q$ from the explicit $I_{ik}(\neg p \wedge \neg q)$, then we can not have $B_i^2\neg p$ and $B_i^3\neg q$, either. This will block many useful deductions from the beginning. ■

6 Conclusion

In this paper, we have proposed a logic integrating the notions of graded trust and belief fusion. By the graded trust in other agents, an agent can induce his partial beliefs from the information he has received. These graded belief can then be merged into his overall belief according to a partial ordering on the set of grades. Two fusion strategies are discussed, leading to the development of respective axiomatic systems. For level cutting strategy, if beliefs at a level are to be discarded, then all those lower than it are also discarded without further examination. On the other hand, for level skipping strategy, only the level under conflict is skipped, and the next level will be considered independent of those discarded before it. Finally, two example are used to illustrate the basic ideas of the logic and the solution to the problem of negative knowledge (or belief) representation is presented.

References

[1] G. Antoniou. *Nonmonotonic Reasoning.* MIT Press, Cambridge, MA, 1997.

[2] L. Åqvist. "Deontic logic". In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Vol II: Extensions of Classical Logic*, pages 605–714. D. Reidel Publishing Company, 1984.

[3] C. Castelfranchi and R. Falcone. "Principle of trust for MAS: cognitive anatomy, social importance, and quantification". In Y. Demazeau, editor, *Proc. of the 3rd International Conference on Multi Agent Systems*, pages 72–79. IEEE, 1998.

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

[5] L. Cholvy. "A logical approach to multi-souces reasoning". In M. Masuch and L. Pólos, editors, *Knowledge Representation and Reasoning under Uncertainty*, LNAI 808, pages 183–196. Springer-Verlag, 1994.

[6] R. Demolombe. "To trust information sources: a proposal for a modal logic framework". In C. Castelfranchi and Y.H. Tan, editors, *Trust and Deception in Virtual Societies.* Kluwer Academic Publisher, 1999.

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

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

[9] J. Hintikka. *Knowledge and Belief.* Cornell University Press, 1962.

[10] C. J. Liau. "A conservative approach to distributed belief fusion". In *Proc. of the Third International Conference on Information Fusion*, pages MoD4–1, 2000.

[11] C.J. Liau. "Logical Systems for Reasoning about Multi-agent Belief, Information Acquisition and Trust". In W. Horn, editor, *Proc. of the 14th European Conference on Artificial Intelligence*, pages 368–372. IOS Presss, 2000.

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

[13] P. Noriega and C. Sierra, editors. *Agent Mediated Electronic Commerce (LNAI 1571).* Springer-Verlag, Berlin, Germany, 1999.

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

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