Trust-based belief change

Emiliano Lorini[1] and **Guifei Jiang**[2] and **Laurent Perrussel**[3]

Abstract. We propose a modal logic that supports reasoning about trust-based belief change. The term *trust-based belief change* refers to belief change that depends on the degree of trust the receiver has in the source of information.

1 Introduction

Trust in information sources plays a crucial role in many domains of interaction between agents, in particular when information sources are either human agents or software agents (e.g., banks, companies, consultants, etc.), typical examples are in the field of e-commerce or in the field of stock and bond market. In the latter case, an agent may receive information from a given source about the evolution of a stock's price. In these situations, the agent's *trust* in the source has an influence on the dynamics of the *belief* about the evolution of the stock's price. The latter belief is fundamental for the agent to decide whether to buy or sell stocks.

The aim of this paper is to improve understanding of the relationship between belief and trust: we propose a logic for reasoning about *trust-based belief change*, that is, belief change that depends on the degree of trust the receiver has in the information source. We call this logic DL-BT which stands for *Dynamic Logic of graded Belief and Trust*. Using this logic, we stress out the interplay between trust and belief change in a modular way. As opposed to numerous approaches such as [4] where the interplay is predefined and thus specific, the logic DL-BT allows to implement different trust-based belief change policies.

On the technical level, the logic DL-BT consists of extending Liau's static modal logic of belief and trust [12] in three different directions: (i) a generalization of Liau's approach to graded trust, (ii) its extension by modal operators of knowledge and by modal operators of graded belief based on Spohn's theory of uncertainty [14], and (iii) by a family of dynamic operators in the style of dynamic epistemic logics (DEL) [16]. The latter allows for the representation of the consequences of a trust-based belief change operation while the second enables to handle iterated belief change.

Our contribution is twofold. First of all, our concept of trust-based belief change does not presuppose that incoming information is necessarily incorporated in the belief set of the agent. This is a key difference with classical belief revision [1] whose primary principle of change (or success postulate) leads any agent to accept new information and to revise her beliefs accordingly. This postulate has been widely criticized in the literature and several approaches of non-prioritized belief revision have been proposed [9]. Credibility-limited revision approach [10, 5] assumes that revision will be successful only if new information is credible, in the sense it does not conflict with the current beliefs of the agent. Differently from this approach, our key criterion for acceptance of new information is not its credibility but trust in the information source.

Secondly, our logic DL-BT provides a solution to the problem of representing the *author* of a communicative act in the DEL-framework. Indeed, existing dynamic epistemic logics [13, 3] do not specify the author of the announcement, as they assume that the announcement is performed by some agent outside the system that is not part of the logic's object language.

The paper is organized as follows. We first present the syntax and the semantics of the logic DL-BT and detail two trust-based belief change policies: an *additive* policy and a *compensatory* policy. The *additive* policy cumulates information received by different information sources. In case different sources provide conflicting information, the *compensatory* policy balances them depending on how much they are trustworthy. We then provide a sound and complete axiomatization for the variant of DL-BT implementing these two policies.

2 Dynamic logic of graded belief and trust

In the next two sections we present the syntax and semantics of the logic DL-BT that combines modal operators of knowledge, graded belief and trust with dynamic operators of trust-based belief change.

2.1 Syntax of DL-BT

Let $Atm = \{p, q, \ldots\}$ be a countable set of propositional atoms and let $Agt = \{i, j, \ldots\}$ be a finite set of agents. Moreover, let $Num = \{0, \ldots, \mathsf{max}\}$ be a finite set of natural numbers with $\mathsf{max} \in \mathbb{N} \setminus \{0\}$ which represents the scale for trust and belief degrees. For instance, the set $Num = \{0, 1, 2, 3, 4, 5\}$ can be interpreted as a *qualitative* scale where 0 stands for 'null' and 5 for 'very high'. Finally, let Plc be a set of trust-based belief change policies.

Let us stress that DL-BT should be conceived as a "family" of logics rather than a single logic, each of which is parameterized by a certain set of trust-based belief change policies Plc. Hereafter, a specific member of the DL-BT family indexed by some set Plc, is denoted DL-BT^{Plc}.

The language $\mathcal{L}$ of DL-BT is defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{K}_i\varphi \mid \mathsf{B}_i^{\geq\alpha}\varphi \mid \mathsf{T}_{i,j}^{\alpha}\varphi \mid [*_i^f\varphi]\varphi$$

where p ranges over Atm, i and j range over Agt, α ranges over $Num\setminus\{0\}$, and f ranges over the set of total functions with domain Agt and codomain Plc.

The other boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$, $\leftrightarrow$ are defined in the standard way. Let Obj be the set of all boolean combinations of atoms in Atm.The elements of Obj are called objective formulas.

K_i is the standard S5 epistemic operator [8]: $\mathsf{K}_i\varphi$ stands for "agent i knows that φ is true".

[1] IRIT-CNRS, University of Toulouse, France, lorini@irit.fr
[2] AIRG-University of Western Sydney, Australia and IRIT-University of Toulouse, France
[3] IRIT, University of Toulouse, France

The formula $\mathsf{B}_i^{\geq\alpha}\varphi$ has to be read "agent i believes that φ is true with strength at least α". Similar operators of graded belief have been studied in the past by [2, 15, 11]. The formula $[*_i^f\varphi]\psi$ has to be read "after agent i has publicly announced that φ is true and each agent j has revised her beliefs according to the trust-based belief change policy $f(j)$, ψ will be true". In other words, the dynamic operator $[*_i^f\varphi]$ allows to represent the effect of agent i's announcement of φ: each agent revises her beliefs according to the trust-based belief change policy prescribed by the function f. Finally, the formula $\mathsf{T}_{i,j}^\alpha\varphi$ has to be read as "agent i trusts agent j's judgement on formula φ with strength α". Note that, when $i = j$, the operator $\mathsf{T}_{i,j}^\alpha$ captures a notion of self-trust (or self-confidence).

We will use the following abbreviations in the rest of the paper. For all $i \in Agt$ and for all $\alpha \in Num \setminus \{0, \mathsf{max}\}$ we define:

$$\widehat{\mathsf{K}}_i\varphi =_{def} \neg\mathsf{K}_i\neg\varphi \qquad \mathsf{B}_i\varphi =_{def} \mathsf{B}_i^{\geq 1}\varphi$$

$$\widehat{\mathsf{B}}_i\varphi =_{def} \neg\mathsf{B}_i\neg\varphi \qquad \mathsf{U}_i\varphi =_{def} \neg\mathsf{B}_i\varphi \wedge \neg\mathsf{B}_i\neg\varphi$$

$$\mathsf{B}_i^{\alpha}\varphi =_{def} \mathsf{B}_i^{\geq\alpha}\varphi \wedge \neg\mathsf{B}_i^{\geq(\alpha+1)}\varphi \qquad \mathsf{B}_i^{\mathsf{max}}\varphi =_{def} \mathsf{B}_i^{\geq\mathsf{max}}\varphi$$

$$\mathsf{B}_i^{0}\varphi =_{def} \neg\mathsf{B}_i\varphi \qquad \mathsf{T}_{i,j}\varphi =_{def} \bigvee_{\alpha\in Num\setminus\{0\}} \mathsf{T}_{i,j}^\alpha\varphi$$

$\widehat{\mathsf{K}}_i$ is the dual of K_i and $\widehat{\mathsf{K}}_i\varphi$ has to be read "φ is compatible with agent i's knowledge".

The operator B_i captures the concept of belief and $\mathsf{B}_i\varphi$ has to be read "agent i believes that φ is true". Indeed, we assume that "believing that φ is true" is the same as "believing that φ is true with strength at least 1".

$\widehat{\mathsf{B}}_i$ is the dual of B_i and $\widehat{\mathsf{B}}_i\varphi$ has to be read "φ is compatible with agent i's beliefs". The operator U_i captures the concept of uncertainty or doubt, and $\mathsf{U}_i\varphi$ has to be read "agent i is uncertain whether φ is true". The operator B_i^α captures the *exact* degree of belief. Specifically, $\mathsf{B}_i^\alpha\varphi$ has to be read "agent i believes that φ is true with strength equal to α". The special case $\mathsf{B}_i^{\mathsf{max}}\varphi$ needs to be defined independently since $\mathsf{B}_i^{\mathsf{max}+1}\varphi$ is not a well-formed formula.

The abbreviation $\mathsf{B}_i^0\varphi$ has to be read "agent i believes that φ with strength 0" which is the same thing as saying that agent i does not believe φ. Finally, $\mathsf{T}_{i,j}\varphi$ has to be read "agent i trusts agent j's judgment on φ".

We call L-BT the static fragment of DL-BT, that is, DL-BT formulas with no dynamic operators $[*_i^f\varphi]$. The language L-BT is defined as follows (as previously, p ranges over Atm, i and j range over Agt and α ranges over $Num\setminus\{0\}$):

$$\varphi ::= p \mid \neg\varphi \mid \varphi\wedge\varphi \mid \mathsf{K}_i\varphi \mid \mathsf{B}_i^{\geq\alpha}\varphi \mid \mathsf{T}_{i,j}^\alpha\varphi$$

2.2 Semantics for the static fragment

Let us first focus on the semantics of the static L-BT formulas. Semantics is defined in terms of possible worlds with a special function for ranking worlds according to their plausibility degrees, and a family of neighbourhood functions for trust.

Definition 1 (Model)**.** *A model is a tuple* $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ *where:*

- *W is a nonempty set of possible worlds or states;*
- *every $\mathcal{E}_i$ is an equivalence relation on W with $\mathcal{E}_i(w) = \{v \in W : w\mathcal{E}_i v\}$ for all $w \in W$;*
- *$\kappa : W \times Agt \to Num$ is a total function mapping each world and each agent to a natural number in Num such that:*

  (Constr1) *for every $w \in W$ and for every $i \in Agt$, there is $v \in W$ such that $w\mathcal{E}_i v$ and $\kappa(v, i) = 0$;*
- *$\mathcal{N}_{i,j} : W \times Num\setminus\{0\} \longrightarrow 2^{2^W}$ is a total function such that for all $w \in W$, for all $i, j \in Agt$, for all $X \subseteq W$ and for all $\alpha, \beta \in Num\setminus\{0\}$:*

  (Constr2) *if $X \in \mathcal{N}_{i,j}(w,\alpha)$ and $\alpha \neq \beta$ then $X \notin \mathcal{N}_{i,j}(w,\beta)$;*

  (Constr3) *if $X \in \mathcal{N}_{i,j}(w,\alpha)$ then $X \in \mathcal{N}_{i,j}(v,\alpha)$ for all $v \in \mathcal{E}_i(w)$;*

  (Constr4) *if $X \in \mathcal{N}_{i,j}(w,\alpha)$ then $X \cap \mathcal{E}_i(w) \neq \emptyset$;*
- *$\mathcal{V} : W \longrightarrow 2^{Atm}$ is a valuation function for propositional atoms.*

As usual, $p \in \mathcal{V}(w)$ means that proposition p is true at world w. The set $\mathcal{E}_i(w)$ is agent i's *information set* at world w: the set of worlds that agent i envisages at world w. As $\mathcal{E}_i$ is an equivalence relation, if $w\mathcal{E}_i v$ then agent i has the same information set at w and v (i.e., agent i has the same knowledge at w and v).

The function κ provides a plausibility grading of the possible worlds for each agent i. $\kappa(w,i) = \alpha$ means that, according to agent i the world w has a degree of exceptionality α or, alternatively, according to agent i the world w has a degree of plausibility $\mathsf{max} - \alpha$. Indeed, following [14], we assume that the degree of plausibility of a world for an agent is the opposite of its exceptionality degree.

(**Constr1**) is a *normality* constraint for the plausibility grading which ensures that an agent can always envisage a world with a minimal degree of exceptionality 0. This constraint is important because it ensures that an agent's beliefs are consistent, e.g., an agent cannot believe φ and $\neg\varphi$ at the same time (see below for more details).

The neighbourhood function $\mathcal{N}_{i,j}$ specifies a trust grading of the subset of possible worlds and is used to interpret the graded trust formulas $\mathsf{T}_{i,j}^\alpha\varphi$. Since each set of possible worlds $X \subseteq W$ is the semantic counterpart of a L-BT formula, the meaning of $X \in \mathcal{N}_{i,j}(w,\alpha)$ is that, at world w, agent i trusts agent j's judgment on the truth of the formula corresponding to X with strength α. **(Constr2)**-**(Constr4)** are natural constraints for trust. Specifically, **(Constr2)** requires that an agent cannot trust the same agent on the same formula with different strengths. **(Constr3)** corresponds to a property of positive introspection for trust, i.e., an agent knows how much she trusts someone. It is worth noting **(Constr3)** and the fact that $\mathcal{E}_i$ is an equivalence relation together imply that if $X \notin \mathcal{N}_{i,j}(w,\alpha)$ then $X \notin \mathcal{N}_{i,j}(v,\alpha)$ for all $v \in \mathcal{E}_i(w)$. The latter corresponds to negative introspection for trust, i.e., if an agent does not trust someone then she knows this. **(Constr4)** claims that an agent's trust in someone must be compatible with her knowledge. Specifically, if agent i trusts agent j's judgement on the truth of some formula, then there should be some world that i envisages in which this formula is true.

We use a neighbourhood semantics for interpreting the graded trust operators $\mathsf{T}_{i,j}^\alpha$ because these modal operators are not normal. We want to allow situations in which, at the same time, agent i trusts agent j's judgement about φ with strength α and i trusts agent j's judgement about $\neg\varphi$ with strength α, without inferring that i trusts agent j's judgement about $\bot$ with strength α, that is, we want formula $\mathsf{T}_{i,j}^\alpha\varphi \wedge \mathsf{T}_{i,j}^\alpha\neg\varphi \wedge \neg\mathsf{T}_{i,j}^\alpha\bot$ to be satisfiable. For example, Bill may trust Mary's judgement about the fact that a certain stock will go upward with strength α (i.e., $\mathsf{T}_{Bill,Mary}^\alpha stockUp$) and, at the same time, trust Mary's judgement about the fact that the stock will not go upward with strength α (i.e., $\mathsf{T}_{Bill,Mary}^\alpha\neg stockUp$), without trusting Mary's judgement about $\bot$ with strength α (i.e., $\neg\mathsf{T}_{Bill,Mary}^\alpha\bot$).[4]

[4] Note that Constraint **(Constr4)** in Definition 1 makes formula $\neg\mathsf{T}_{i,j}^\alpha\bot$ valid for every trust value α. Thus, if $\mathsf{T}_{i,j}^\alpha$ was a normal modal operator, $\neg(\mathsf{T}_{i,j}^\alpha\varphi \wedge \mathsf{T}_{i,j}^\alpha\neg\varphi)$ would have been valid, which is highly counterintuitive.

Before providing truth conditions of L-BT formulas, we follow [14] and lift the exceptionality of a possible world to the exceptionality of a formula viewed as a set of worlds.

Definition 2 (Exceptionality of a formula)**.** *Let $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ be a model. Moreover, let $\|\varphi\|_{w,i} = \{v \in W : v \in \mathcal{E}_i(w)$ and $M, v \models \varphi\}$ be the set of worlds envisaged by agent i at w in which φ is true. The exceptionality degree of formula φ for agent i at world w, denoted by $\kappa_{w,i}(\varphi)$, is defined as follows:*

$$\kappa_{w,i}(\varphi) = \begin{cases} \min_{v\in\|\varphi\|_{w,i}} \kappa(v, i) & \text{if } \|\varphi\|_{w,i} \neq \emptyset \\ \mathsf{max} & \text{if } \|\varphi\|_{w,i} = \emptyset \end{cases}$$

The exceptionality degree of a formula φ captures the extent to which φ is considered to be exceptional by the agent. The value $\kappa_{w,i}(\neg\varphi)$ corresponds to the *degree of necessity* of φ according to agent i at w, in the sense of possibility theory [7]. The following definition provides truth conditions for L-BT formulas.

Definition 3 (Truth conditions)**.** *Let $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ be a model and let $w \in W$. Then:*

$$\begin{array}{rll} M, w \models p & \textit{iff} & p \in \mathcal{V}(w) \\ M, w \models \neg\varphi & \textit{iff} & M, w \not\models \varphi \\ M, w \models \varphi \wedge \psi & \textit{iff} & M, w \models \varphi \text{ and } M, w \models \psi \\ M, w \models \mathsf{K}_i\varphi & \textit{iff} & \forall v \in \mathcal{E}_i(w) : M, v \models \varphi \\ M, w \models \mathsf{B}_i^{\geq\alpha}\varphi & \textit{iff} & \kappa_{w,i}(\neg\varphi) \geq \alpha \\ M, w \models \mathsf{T}_{i,j}^{\alpha}\varphi & \textit{iff} & \|\varphi\|_M \in \mathcal{N}_{i,j}(w, \alpha) \text{ with} \\ & & \|\varphi\|_M = \{v \in W : M, v \models \varphi\} \end{array}$$

In the following, we say that the L-BT formula φ is valid, denoted by $\models \varphi$, if for every model M and for every world w in M we have $M, w \models \varphi$. Moreover, we say that φ is satisfiable if $\neg\varphi$ is not valid. The following validity highlights that beliefs are necessarily consistent: $\models \neg(\mathsf{B}_i\varphi \wedge \mathsf{B}_i\neg\varphi)$. In the next section, we provide truth conditions for DL-BT formulas $[*_i^f\varphi]\psi$, after introducing the concept of trust-based belief change policy.

2.3 Trust-based belief change policies

A trust-based belief change policy specifies the way an agent's plausibility ranking of possible worlds should be modified depending on the agent's trust in the information source.

2.3.1 Additive policy

We start by considering an *additive* trust-based belief change policy and denote it by the symbol add. It is inspired by Darwiche & Pearl's well-known iterated belief revision method [6].

Definition 4 (Additive policy)**.** *Let $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ be a model and let f be a function with domain Agt and codomain Plc such that $f(j) = add$. Then, for all $w \in W$, we define:*

$$\kappa^{*_i^f\varphi}(w, j) = \begin{cases} \textit{Case 1.} & \kappa(w, j) \\ & \text{if } M, w \models \neg\mathsf{T}_{j,i}\varphi \\ \textit{Case 2.} & \kappa(w, j) - \kappa_{w,j}(\varphi) \\ & \text{if } M, w \models \varphi \wedge \mathsf{T}_{j,i}\varphi \\ \textit{Case 3.} & Cut(\alpha + \kappa(w, j)) \\ & \text{if } M, w \models \neg\varphi \wedge \mathsf{T}_{j,i}^{\alpha}\varphi \end{cases}$$

where:

$$Cut(x) = \begin{cases} x & \text{if } 0 \leq x \leq \mathsf{max} \\ \mathsf{max} & \text{if } x > \mathsf{max} \\ 0 & \text{if } x < 0 \end{cases}$$

Suppose that the information source i publicly announces that φ is true. Then, according to Definition 4, the additive rule rigidly boosts the $\neg\varphi$-worlds up from where they currently are by the degree of trust agent j has in the information source i. We show below that this policy guarantees that information received by different information sources is cumulated, in the sense that agent j 'does not forget' her previous degree of belief about φ.

Note that Case 3 is well-defined because of Constraint (**Constr2**), agent j cannot trust agent i with different strengths. Function Cut is a minor technical device, taken from [2], which ensures that the new plausibility assignment fits into the finite set of natural numbers Num. Moreover, note that the situation in which agent j knows that φ is false is a special case of the preceding Case 1. Indeed, because of (**Constr4**) in Definition 1, formula $\mathsf{K}_j\neg\varphi \rightarrow \neg\mathsf{T}_{j,i}\varphi$ is valid. Consequently, if i's announcement of φ is incompatible with j's knowledge, then i's announcement of φ does not have any effect on j's beliefs.

We are in position to give the truth condition of the dynamic operator $[*_i^f\varphi]$ for the logic DL-BT$^{\{add\}}$.

Definition 5 (Truth conditions (cont.))**.** *Let $Plc = \{add\}$, let $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ be a model and $w \in W$. Then:*

$$M, w \models [*_i^f\varphi]\psi \quad \textit{iff} \quad M^{*_i^f\varphi}, w \models \psi$$

where $M^{*_i^f\varphi} = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa^{*_i^f\varphi}, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ and function $\kappa^{*_i^f\varphi}$ is defined according to the preceding Definition 4.

We generalize the notions of validity and satisfiability for DL-BT$^{\{add\}}$ formulas from the the notions of validity and satisfiability for L-BT formulas in the obvious way.

2.3.2 Properties of the additive policy

The next proposition highlights that the additive policy defined above is syntax independent, in the sense that two public announcements with logically equivalent formulas produce the same effects. This is a consequence of the fact that the graded trust operator $\mathsf{T}_{i,j}^{\alpha}$ is closed under logical equivalence.

Proposition 1. *If $f(j) = add$ for all $j \in Agt$ and $\models \varphi_1 \leftrightarrow \varphi_2$ then:*

$$\models [*_i^f\varphi_1]\psi \leftrightarrow [*_i^f\varphi_2]\psi \tag{1}$$

The following proposition captures two fundamental properties of the additive policy.

Proposition 2. *For all $i, j \in Agt$ and for all $\alpha \in Num \setminus \{0\}$, if $f(j) = add$ and $\psi \in Obj$ then:*

$$\models (\mathsf{B}_j^{\alpha}\psi \wedge \neg\mathsf{T}_{j,i}\varphi) \rightarrow [*_i^f\varphi]\mathsf{B}_j^{\alpha}\psi \tag{2}$$

$$\models \mathsf{T}_{j,i}\psi \rightarrow [*_i^f\psi]\mathsf{B}_j\psi \tag{3}$$

According to validity (2), if an agent does not trust the information source then her beliefs about objective facts are not affected by what the information source announces. The validity (3) is a weakening of the AGM success postulate: agent j will revise her beliefs with the objective formula ψ only if j trusts the information source's judgment on ψ. Validity (3) can be generalized to a sequence of announcements of any length n.

Proposition 3. *For all $j, i_1, \ldots, i_n \in Agt$ and for all $n \in \mathbb{N}$, if $f(j) = add$ and $\psi_n \in Obj$ then:*

$$\models \mathsf{T}_{j,i_n}\psi_n \to [*^{f_1}_{i_1}\psi_1] \ldots [*^{f_n}_{i_n}\psi_n]\mathsf{B}_j\psi_n \quad (4)$$

Let us consider the special case of the preceding validity with $n = 2$ and $\psi \in Obj$. We have:

$$\models \mathsf{T}_{j,i_2}\psi \to [*^{f_1}_{i_1}\neg\psi][*^{f_n}_{i_2}\psi]\mathsf{B}_j\psi. \quad (5)$$

This means that, in the case of the additive policy, if two sources provide contradictory information, then the receiver will give priority to the last information source, if she trusts her. Let us now illustrate the cumulative effect of the additive policy.

Proposition 4. *For all $i, j \in Agt$ and for all $\alpha, \beta \in Num \setminus \{0\}$, if $f(j) = add$ and $\varphi \in Obj$ then:*

$$\models (\mathsf{T}^{\alpha}_{j,i}\varphi \wedge \mathsf{B}^{\beta}_{j}\varphi) \to [*^{f}_{i}\varphi]\mathsf{B}^{Cut(\alpha+\beta)}_{j}\varphi \quad (6)$$

$$\models (\mathsf{T}^{\alpha}_{j,i}\varphi \wedge \neg\mathsf{B}_{j}\varphi) \to [*^{f}_{i}\varphi]\mathsf{B}^{\alpha}_{j}\varphi \quad (7)$$

Validity (6) highlights that the additive policy takes into account not only agent j's trust in the source, but also what agent j believed before the source's announcement. In particular, if agent j trusts i's judgment on the objective formula φ with degree α and believes φ with strength β then, after i's announcement of φ, j will believe φ with strength $Cut(\alpha + \beta)$. Validity (7) captures the complementary case in which agent j does not believe φ before the announcement. In this case, the strength of j's belief about φ is only determined by j's trust in the information source i.

The two validities of Proposition 4 can actually be generalized to a sequence of announcements of any length n as it is highlighted by Proposition 5. In particular, (i) if φ is an objective formula and j believes φ with a certain degree α, then j's degree of belief about φ at the end of a sequence of n announcements of φ is equal to the sum of α and j's degrees of trust in the sources of the announcements; (i) if φ is an objective formula and j does not believe φ, then j's degree of belief about φ at the end of a sequence of n announcements of φ is equal to the sum of j's degrees of trust in the sources of the announcements. More generally, the additive policy cumulates information about objective facts coming from different sources.

Proposition 5. *For all $j, i_1, \ldots, i_n \in Agt$, for all $\alpha_1 \ldots, \alpha_n, \gamma \in Num \setminus \{0\}$ and for all $n \in \mathbb{N}$, if $f_1(j) = \ldots = f_n(j) = add$ and $\varphi \in Obj$ then:*

$$\models (\mathsf{T}^{\alpha_1}_{j,i_1}\varphi \wedge \ldots \wedge \mathsf{T}^{\alpha_n}_{j,i_n}\varphi \wedge \mathsf{B}^{\gamma}_{j}\varphi) \to [*^{f_1}_{i_1}\varphi] \ldots [*^{f_n}_{i_n}\varphi]\mathsf{B}^{Cut(\alpha_1+\ldots+\alpha_n+\gamma)}_{i}\varphi \quad (8)$$

$$\models (\mathsf{T}^{\alpha_1}_{j,i_1}\varphi \wedge \ldots \wedge \mathsf{T}^{\alpha_n}_{j,i_n}\varphi \wedge \neg\mathsf{B}_{j}\varphi) \to [*^{f_1}_{i_1}\varphi] \ldots [*^{f_n}_{i_n}\varphi]\mathsf{B}^{Cut(\alpha_1+\ldots+\alpha_n)}_{i}\varphi \quad (9)$$

In propositions 2–5, we only consider objective formulas as they do not hold in general. If we drop the restriction to objective formulas, the validity (3) in Proposition 2 does not work anymore. To see this, suppose that ψ is a Moore-like sentence of the form $p \wedge \neg\mathsf{B}_j p$. Then, the formula $\mathsf{T}_{j,i}(p \wedge \neg\mathsf{B}_j p) \to [*^{f}_{i}(p \wedge \neg\mathsf{B}_j p)]\mathsf{B}_j(p \wedge \neg\mathsf{B}_j p)$ is clearly not valid. In fact, $\mathsf{B}_j(p \wedge \neg\mathsf{B}_j p)$ is equivalent to $\perp$. Similar observations hold for Propositions 3–5.

Notice that the additive policy satisfies the following commutativity property.

Proposition 6. *For all $i_1, i_2 \in Agt$, if $f(j) = add$ for all $j \in Agt$ and $\varphi \in Obj$ then:*

$$\models [*^{f}_{i_1}\varphi][*^{f}_{i_2}\varphi]\psi \leftrightarrow [*^{f}_{i_2}\varphi][*^{f}_{i_1}\varphi]\psi \quad (10)$$

This means that if all agents adopt the additive policy then the order of the announcements of an objective formula φ performed by several information sources does not matter.

Example 1. *Let us illustrate the additive policy. Assume that $Num = \{0, 1, 2, 3, 4, 5\}$ s.t. 0 means 'null', 1 means 'very weak', 2 means 'weak', 3 means 'fair', 4 means 'strong' and 5 means 'very strong'.*

Bill has to decide whether he buys a certain stock. He hesitates because he is uncertain whether the stock will go upward ($stockUp$). Assume the following initial epistemic state for Bill:

$$Hyp1 =_{def} \mathsf{U}_{Bill}\, stockUp$$

Bill asks two stockbrokers their opinions: Mary and Jack. He first asks Mary. Then, he asks Jack. Both Mary and Jack say that the stock will go upward and that it is convenient to buy it. We assume that Bill trusts fairly *Mary's judgement on $stockUp$, and Bill trusts* very weakly *Jack's judgement on $stockUp$:*

$$Hyp2 =_{def} \mathsf{T}^{3}_{Bill,Mary}\, stockUp \wedge \mathsf{T}^{1}_{Bill,Jack}\, stockUp$$

Suppose Bill uses the additive policy. In this situation, after having received the information from Mary and Jack, Bill will strongly believe that proposition $stockUp$ is true. As Proposition 5 above highlighs, Bill cumulates the information provided by the two information sources. Specifically, if $f(Bill) = f'(Bill) = add$ then:

$$\models (Hyp1 \wedge Hyp2) \to [*^{f}_{Mary}\, stockUp][*^{f'}_{Jack}\, stockUp]\mathsf{B}^{4}_{Bill}\, stockUp.$$

Now, suppose that Mary and Jack provide contradictory information about proposition $stockUp$. As highlighted by Proposition 3, priority will be given to the last information source. That is, if $f(Bill) = f'(Bill) = add$ then:

$$\models (Hyp1 \wedge Hyp2) \to [*^{f}_{Mary}\neg stockUp][*^{f'}_{Jack}\, stockUp]\mathsf{B}_{Bill}\, stockUp.$$

In the next section we present a new policy, the *compensatory* policy that does not satisfy the general property given in Proposition 3. In case of two contradictory information provided by two sources, the *compensatory* policy balances them depending on the degrees of trust in the sources.

2.3.3 Compensatory policy

The compensatory policy, denoted by the symbol $comp$, is defined as follows.

Definition 6 (Compensatory policy)**.** *Let $M = (W, \{\mathcal{E}_i\}_{i\in Agt}, \kappa, \{\mathcal{N}_{i,j}\}_{i,j\in Agt}, \mathcal{V})$ be a model and let f be a function with domain Agt and codomain Plc such that $f(j) = comp$. Then, for all $w \in W$, we define:*

$$\kappa^{*^{f}_{i}\varphi}(w,j) = \begin{cases} \textit{Case 1.} & \kappa(w,j) \\ & \textit{if } M, w \models \neg\mathsf{T}_{j,i}\varphi \\ \textit{Case 2.} & Cut(\kappa(w,j) - \alpha) \\ & \textit{if } M, w \models \varphi \wedge \mathsf{T}^{\alpha}_{j,i}\varphi \\ \textit{Case 3.} & Cut(\alpha + \kappa(w,j)) \\ & \textit{if } M, w \models \neg\varphi \wedge \mathsf{T}^{\alpha}_{j,i}\varphi \wedge \widehat{\mathsf{B}}_j\varphi \\ \textit{Case 4.} & \kappa(w,j) \\ & \textit{if } M, w \models \neg\varphi \wedge \mathsf{T}_{j,i}\varphi \wedge \mathsf{B}_j\neg\varphi \end{cases}$$

where $Cut(x)$ is the same as in Definition 4.

Let us focus on Cases 2, 3 and 4, as Case 1 is the same as the one in Definition 4. Case 2 states that, if j trusts i's judgment on φ then, after i's announcement of φ, the exceptionality degree of a φ-world for j should be decreased depending on how much j trusts i, in order to decrease the strength of j's belief about $\neg\varphi$. Cases 3 and 4 distinguish the situation in which φ is compatible with j's beliefs from the situation in which it is not. For Case 3, the exceptionality degree of a $\neg\varphi$-world for j should be increased in order to increase the strength of j's belief about φ. For Case 4, agent j should not change her plausibility ordering in order to preserve consistency of beliefs. Case 4 guarantees that **(Constr1)** in Definition 1 is preserved.

The truth condition of the dynamic operator $[*_i^f \psi]$ as well as the notion of validity for the logic DL-BT$^{\{add,comp\}}$ are defined in a similar way from the ones for the logic DL-BT$^{\{add\}}$ given above.

It is important to remark that the compensatory policy as well as the additive policy guarantee that the updated model $M^{*_i^f \varphi}$ is indeed a model in the sense of Definition 1. In particular:

Proposition 7. *If M is a model in the sense of Definition 1 and $f(j) \in \{add, comp\}$ for all $j \in Agt$, then $M^{*_i^f \varphi}$ is a model in the sense of Definition 1 too.*

2.3.4 Properties of the compensatory policy

Let us consider some basic properties of the compensatory policy. The first point to remark is that, different from the additive policy, the compensatory policy does not satisfy the weakening of the success postulate of Proposition 2. That is, if $f(j) = comp$ and $\psi \in Obj$, the formula $\mathsf{T}_{j,i}\psi \rightarrow [*_i^f \psi]\mathsf{B}_j\psi$ is not valid.

The following Proposition 8 provides a list of validities for the compensatory policy.

Proposition 8. *For all $i, j, i_1, i_2 \in Agt$, for all $\alpha, \alpha_1, \alpha_2, \beta \in Num \setminus \{0\}$, if $f(j) = f'(j) = comp$ and $\varphi \in Obj$ then:*

$$\models (\mathsf{T}_{j,i}^{\alpha}\varphi \wedge \mathsf{B}_j^{\beta}\varphi) \rightarrow [*_i^f \varphi]\mathsf{B}_j^{Cut(\alpha+\beta)}\varphi \quad (11)$$

$$\models (\mathsf{T}_{j,i}^{\alpha}\varphi \wedge \mathsf{U}_j\varphi) \rightarrow [*_i^f \varphi]\mathsf{B}_j^{\alpha}\varphi \quad (12)$$

$$\models (\mathsf{T}_{j,i}^{\alpha}\neg\varphi \wedge \mathsf{B}_j^{\beta}\varphi) \rightarrow [*_i^f \neg\varphi]\mathsf{B}_j^{Cut(\beta-\alpha)}\varphi \quad (13)$$

$$\models (\mathsf{T}_{j,i_1}^{\alpha_1}\varphi \wedge \mathsf{T}_{j,i_2}^{\alpha_2}\neg\varphi \wedge \mathsf{B}_j^{\beta}\varphi) \rightarrow [*_{i_1}^f \varphi][*_{i_2}^{f'} \neg\varphi]\mathsf{B}_j^{Cut(Cut(\beta+\alpha_1)-\alpha_2)}\varphi \quad (14)$$

$$\models (\mathsf{T}_{j,i_1}^{\alpha_1}\varphi \wedge \mathsf{T}_{j,i_2}^{\alpha_2}\neg\varphi \wedge \mathsf{U}_j\varphi) \rightarrow [*_{i_1}^f \varphi][*_{i_2}^{f'} \neg\varphi]\mathsf{B}_j^{Cut(\alpha_1-\alpha_2)}\varphi \quad (15)$$

According to the validities (11) and (12), if i announces the objective formula φ, φ is consistent with j's beliefs and j adopts the compensatory policy, then j will increase the strength of her belief about φ w.r.t. her trust's degree in i. The two validities distinguish the situation whether j believes φ, or whether j is uncertain about φ. The last three key validities (13)–(15) characterize the compensatory aspect. According to the validity (13), if φ is an objective formula then, after i announces that φ is false, j will decrease the strength of her belief about φ depending on how much she trusts i. The validities (14) and (15) considers information about objective facts coming from two different sources. Suppose that i_1 says φ, while i_2 says $\neg\varphi$. Then, j should compensate the information received from i_1 by decreasing the strength of her belief about φ depending on how much she trusts i_2. Let us illustrate this by revisiting our previous example.

Example 2. *Let us suppose that Mary and Jack provide contradictory information about proposition $stockUp$. Suppose Bill trusts Mary's judgment on $stockUp$ with degree 3 and trusts Jack's judgment on $\neg stockUp$ with degree 1:*

$$Hyp2' =_{def} \mathsf{T}_{Bill,Mary}^{3} stockUp \wedge \mathsf{T}_{Bill,Jack}^{1} \neg stockUp$$

Now, assume Mary announces that $stockUp$ is true and Jack announces that $stockUp$ is false. If Bill adopts the compensatory policy, he will then believe that $stockUp$ is true with strength $3 - 1 = 2$. That is, if $f(Bill), f'(Bill) = \{comp\}$:

$$\models (Hyp1 \wedge Hyp2') \rightarrow [*_{Mary}^f stockUp][*_{Jack}^{f'} \neg stockUp]\mathsf{B}_{Bill}^{2} stockUp.$$

3 Axiomatization

In this section, we provide a complete axiomatization for the variant of DL-BT where $Plc = \{add, comp\}$, namely DL-BT$^{\{add,comp\}}$. This logic has so-called reduction axioms which allow to reduce every formula to an equivalent L-BT formula without dynamic operators $[*_j^f \psi]$. That elimination together with the rule of replacement of equivalent axioms and rules of inference for the static logic L-BT provides an axiomatization.

Proposition 9 provides reduction axioms for boolean formulas, as well as the knowledge and graded trust operators.

Proposition 9. *The following equivalences are valid:*

$$\begin{array}{lll}
[*_j^f \varphi]p & \leftrightarrow & p \\
[*_j^f \varphi]\neg\psi & \leftrightarrow & \neg[*_j^f \varphi]\psi \\
[*_j^f \varphi](\psi_1 \wedge \psi_2) & \leftrightarrow & ([*_j^f \varphi]\psi_1 \wedge [*_j^f \varphi]\psi_2) \\
[*_j^f \varphi]\mathsf{K}_i\psi & \leftrightarrow & \mathsf{K}_i[*_j^f \varphi]\psi \\
[*_j^f \varphi]\mathsf{T}_{i,k}^{\alpha}\psi & \leftrightarrow & \mathsf{T}_{i,k}^{\alpha}[*_j^f \varphi]\psi
\end{array}$$

The following abbreviation is useful to formulate the reduction axioms for the graded belief operators. For all $\alpha > \mathsf{max}$ we give the following abbreviation:

$$\mathsf{B}_i^{\geq\alpha}\varphi =_{def} \mathsf{K}_i\varphi.$$

Proposition 10 provides the reduction axiom for the graded belief operators based on the additive policy.

Proposition 10. *Let $f(i) = add$ and $\alpha \in Num \setminus \{0\}$. Then, the following equivalence is valid:*

$$[*_j^f \varphi]\mathsf{B}_i^{\geq\alpha}\psi \leftrightarrow \Big((\neg\mathsf{T}_{i,j}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}[*_j^f \varphi]\psi) \wedge \bigwedge_{\beta\in Num\setminus\{0\}, \gamma_1\in Num} \big((\mathsf{T}_{i,j}^{\beta}\varphi \wedge \mathsf{B}_i^{\gamma_1}\neg\varphi) \rightarrow (\mathsf{B}_i^{\geq\alpha+\gamma_1}(\varphi \rightarrow [*_j^f \varphi]\psi) \wedge \mathsf{B}_i^{\geq Cut(\alpha-\beta)}(\neg\varphi \rightarrow [*_j^f \varphi]\psi))\big)\Big)$$

Proposition 11 provides the reduction axiom for the graded belief operators based on the compensatory policy.

Proposition 11. *Let $f(i) = comp$ and $\alpha \in Num \setminus \{0\}$. Then, the following equivalence is valid:*

$$[*_j^f \varphi]\mathsf{B}_i^{\geq\alpha}\psi \leftrightarrow \Big((\neg\mathsf{T}_{i,j}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}[*_j^f \varphi]\psi) \wedge \bigwedge_{\beta\in Num\setminus\{0\}} \big(((\mathsf{T}_{i,j}^{\beta}\varphi \wedge \widehat{\mathsf{B}}_i\varphi) \rightarrow (\mathsf{B}_i^{\geq\alpha+\beta}(\varphi \rightarrow [*_j^f \varphi]\psi) \wedge \mathsf{B}_i^{\geq Cut(\alpha-\beta)}(\neg\varphi \rightarrow [*_j^f \varphi]\psi))) \wedge ((\mathsf{T}_{i,j}^{\beta}\varphi \wedge \mathsf{B}_i\neg\varphi) \rightarrow (\mathsf{B}_i^{\geq\alpha+\beta}(\varphi \rightarrow [*_j^f \varphi]\psi) \wedge \mathsf{B}_i^{\geq\alpha}(\neg\varphi \rightarrow [*_j^f \varphi]\psi))))\Big)$$

These two propositions translate the different cases considered in Definitions 4 and 6. For instance, line 1 of Prop.10 describes Case 1 of Def. 4 (no change if no trust) while lines 2–4 correspond to the two options for changing when trust holds (Cases 2 and 3 of Def. 4).

As the rule of replacement of equivalences $\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]}$ preserves validity, the equivalences of Propositions 9, 10 and 11 together with this allow to reduce every DL-BT$^{\{add,comp\}}$ formula to an equivalent L-BT formula. Call τ the mapping which iteratively applies the above equivalences from the left to the right, starting from one of the innermost modal operators. τ pushes the dynamic operators inside the formula, and finally eliminates them when facing an atomic formula.

Proposition 12. *Let φ be a DL-BT$^{\{add,comp\}}$ formula. Then: (i) $\tau(\varphi)$ has no dynamic operators $[*_j^f \psi]$, and (ii) $\tau(\varphi) \leftrightarrow \varphi$ is valid.*

The axiomatic system of the logic DL-BT$^{\{add,comp\}}$ consists of the axioms and rules of inference in Figure 1. Notice that the rule of necessitation for graded belief (i.e., from φ infer $\mathsf{B}_i^{\geq\alpha}\varphi$) does not need to be added, as it is deducible from the rule of necessitation for knowledge, the fifth axiom for graded belief and modus ponens.

(1) **Axioms for** L-BT :

- all tautologies of classical propositional logic
- axioms K, T, 4 and 5 for knowledge
  * $\mathsf{K}_i(\varphi \rightarrow \psi) \rightarrow (\mathsf{K}_i\varphi \rightarrow \mathsf{K}_i\psi)$
  * $\mathsf{K}_i\varphi \rightarrow \varphi$
  * $\mathsf{K}_i\varphi \rightarrow \mathsf{K}_i\mathsf{K}_i\varphi$
  * $\neg\mathsf{K}_i\varphi \rightarrow \mathsf{K}_i\neg\mathsf{K}_i\varphi$
- axioms for graded belief
  * $\mathsf{B}_i^{\geq\alpha}(\varphi \rightarrow \psi) \rightarrow (\mathsf{B}_i^{\geq\alpha}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}\psi)$
  * $\neg(\mathsf{B}_i^{\geq 1}\varphi \wedge \mathsf{B}_i^{\geq 1}\neg\varphi)$
  * $\mathsf{B}_i^{\geq\alpha}\varphi \rightarrow \mathsf{K}_i\mathsf{B}_i^{\geq\alpha}\varphi$
  * $\neg\mathsf{B}_i^{\geq\alpha}\varphi \rightarrow \mathsf{K}_i\neg\mathsf{B}_i^{\geq\alpha}\varphi$
  * $\mathsf{K}_i\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}\varphi$
  * $\mathsf{B}_i^{\geq\alpha+1}\varphi \rightarrow \mathsf{B}_i^{\geq\alpha}\varphi$
- axioms for graded trust
  * $\mathsf{T}_{i,j}^{\alpha}\varphi \rightarrow \neg\mathsf{T}_{i,j}^{\beta}\varphi$ if $\alpha \neq \beta$
  * $\mathsf{T}_{i,j}^{\alpha}\varphi \rightarrow \mathsf{K}_i\mathsf{T}_{i,j}^{\alpha}\varphi$
  * $\mathsf{T}_{i,j}^{\alpha}\varphi \rightarrow \widehat{\mathsf{K}}_i\varphi$

(2) **Rules of inference for** L-BT:

- from φ and $\varphi \rightarrow \psi$ infer ψ
- from φ infer $\mathsf{K}_i\varphi$
- from $\varphi \leftrightarrow \psi$ infer $\mathsf{T}_{i,j}^{\alpha}\varphi \leftrightarrow \mathsf{T}_{i,j}^{\alpha}\psi$

(3) **Further axioms and rules of inference for** DL-BT:

- from $\psi_1 \leftrightarrow \psi_2$ infer $\varphi \leftrightarrow \varphi[\psi_1/\psi_2]$
- reduction axioms of Propositions 9, 10 and 11

Figure 1. Axiomatization of DL-BT$^{\{add,comp\}}$

Theorem 1. *The logic DL-BT$^{\{add,comp\}}$ is completely axiomatized by the principles given in Figure 1.*

(Sketch). Thanks to Prop. 12 and the fact that DL-BT$^{\{add,comp\}}$ is a conservative extension of L-BT, we only need to prove that L-BT is completely axiomatized by the group of axioms (1) and the group of rules of inference (2) in Figure 1. The proof consists of two steps. First, we provide a relational semantics for L-BT and prove that this semantics is equivalent to the L-BT semantics of Def. 1. Then, we use the canonical model construction in order to show that the group of axioms (1) and the group of rules of inference (2) in Fig. 1 provide a complete axiomatics for L-BT with respect to this semantics. □

4 Conclusion

We have proposed a dynamic logic of graded belief and trust that supports reasoning about trust-based belief change. We have considered two kinds of trust-based belief change policy and studied their logical properties in detail. In addition, we have provided a sound and complete axiomatization for our logic.

Following the belief revision tradition, in future work we plan to extend the present work with an axiomatic analysis of the additive and compensatory policies. More concretely, for every policy, we intend to come up with a list of postulates that fully characterize it.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the French ANR project EmoTES Emotions in strategic interaction: theory, experiments, logical and computational studies, contract No. 11-EMCO-004-01.

REFERENCES

[1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson, 'On the logic of theory change: partial meet contraction and revision functions', *Journal of Symbolic Logic*, **50**(2), 510–530, (1985).

[2] G. Aucher, 'A combined system for update logic and belief revision', in *Proc. of PRIMA 2004*, volume 3371 of *LNAI*, pp. 1–18, (2005).

[3] A. Baltag, L. Moss, and S. Solecki, 'The logic of public announcements, common knowledge and private suspicions', in *Proc. of TARK'98*, pp. 43–56, San Francisco, CA, (1998). Morgan Kaufmann.

[4] S. Barber and J. Kim, 'Belief revision process based on trust: Agents evaluating reputation of information sources', in *Proc. of the Workshop on Deception, Fraud, and Trust in Agent Societies*, pp. 73–82. Springer-Verlag, (2001).

[5] R. Booth, E. Fermé, S. Konieczny, and R. P. Pérez, 'Credibility-limited revision operators in propositional logic', in *Proceedings of KR'12*, pp. 116–125. AAAI Press, (2012).

[6] A. Darwiche and J. Pearl, 'On the logic of iterated belief revision', *Artificial Intelligence*, **89**, 1–29, (1996).

[7] D. Dubois and H. Prade, 'Possibility theory: qualitative and quantitative aspects', in *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, 169–226, Kluwer, (1998).

[8] R. Fagin, J. Halpern, and Y. Moses, *Reasoning About Knowledge*, MIT Press, 1995.

[9] S. O. Hansson, 'A survey of non-prioritized belief revision', *Erkenntnis*, **50**, 413–427, (1999).

[10] S. O. Hansson, E. Fermé, J. Cantwell, and M. Falappa, 'Credibility limited revision', *Journal of Symbolic Logic*, **66**, 1581–1596, (2001).

[11] N. Laverny and J. Lang, 'From knowledge-based programs to graded belief-based programs, part ii: off-line reasoning', in *Proc. of IJCAI'05*, pp. 497–502, (2005).

[12] C.-J. Liau, 'Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation', *Art. Intel.*, **149**(1), 31–60, (2003).

[13] J. A. Plaza, 'Logics of public communications', in *Proc. of the 4th International Symposium on Methodologies for Intelligent Systems*, 201-216, (1989).

[14] W. Spohn, 'Ordinal conditional functions: a dynamic theory of epistemic states', in *Causation in decision, belief change and statistics*, 105–134, Kluwer, (1988).

[15] H. van Ditmarsch, 'Prolegomena to dynamic logic for belief revision', *Synthese*, **147**(2), 229–275, (2005).

[16] H. van Ditmarsch, W. van der Hoek, and B. Kooi, *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*, Springer, 2007.