Reasoning About Belief, Evidence and Trust in a Multi-agent Setting

Fenrong Liu[1] and Emiliano Lorini[2(✉)]

[1] Tsinghua University, Beijing, China
[2] IRIT-CNRS, Toulouse University, Toulouse, France
lorini@irit.fr

Abstract. We present a logic for reasoning about the interplay between belief, evidence and trust in a multi-agent setting. We call this logic DL-BET which stands for "Dynamic Logic of Belief, Evidence and Trust". According to DL-BET, if the amount of evidence in support a given fact φ and the ratio of evidence in support of φ to the total amount of evidence in support of either φ or its negation are sufficient then, as a consequence, one should be willing to believe φ. We provide a sound and complete axiomatization for the logic and illustrate its expressive power with the aid of a concrete example.

1 Introduction

As emphasized by the philosopher A.J. Ayer, the connection between evidence and belief is an essential aspect of human rationality:

> "...A rational man is one who makes a proper use of reason: and this implies, among other things, that he correctly estimates the strength of evidence" [3, p. 3].

The connection between belief and evidence is also relevant for artificial intelligence (AI) and, in particular, for reasoning under uncertainty [32] and information fusion [17]. For instance, information fusion can be conceived as an aggregation process which aims to extract truthful knowledge from incomplete or uncertain information coming from various sources of evidence.

The aim of the present paper is to provide a logic of the interplay between evidence and trust, on the one hand, and between evidence and belief, on the other hand. We call this logic DL-BET which stands for "Dynamic Logic of Belief, Evidence and Trust". Specifically, DL-BET supports reasoning about an agent's belief formation and belief change due to evidence provided by reliable information sources.

The central idea of DL-BET is that an agent accumulates *evidence* in support of a given fact φ from other agents in the society and the body of evidence in support of φ can become a reason to *believe* φ. *Trust* is a necessary condition for an agent to accept the information provided by another agent and to integrate it as a new piece of evidence in support of a given fact.

B. An et al. (Eds.): PRIMA 2017, LNAI 10621, pp. 71–89, 2017.
https://doi.org/10.1007/978-3-319-69131-2_5

A central assumption of the logic DL-BET is that, to form a belief that a certain fact φ is true, an agent is sensitive both (i) to the *amount* of evidence in support of φ, and (ii) to the *ratio* of evidence in support of φ to the total amount of evidence in support of either φ or its negation. The notion of "amount of evidence" is reminiscient of Keynes's well-known concept of "weight of argument", as clearly defined in the following paragraph from the famous treatise on probability:

> "...As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according as the new knowledge strengthens the unfavourable or the favourable evidence; but something seems to have increased in either case, - we have more substantial basis upon which to rest our conclusion. I express this by saying that an accession of new evidence increases the weight of argument" [25, p. 77].

The present work is mainly theoretical but, we believe, it can offer interesting insights for people working on multi-agent system (MAS) applications in which agents are supposed to be artificial entities such as a robot, a chatbot or a conversational agent interacting with a human user. Such agents may be endowed with the capability of forming beliefs on the basis of the collected evidence. By way of example, consider a chatbot similar to Apple's Siri or Microsoft's Cortana connected to the Internet who has to provide information to the human user about the quality of a certain movie. In particular, the human user wants to know whether a certain movie is good or bad. The chatbot has access to different recommendation systems about movies in the Internet which are more or less reliable (e.g., Netflix, Rotten Tomatoes, IMDb). The chatbot will form the belief that the movie is good and inform the human user about this, depending on the evidence it possesses in support of this fact.

The paper is organized as follows. In Sect. 2 we present the syntax and semantics of the logic DL-BET and discuss some of its general properties. The semantics of DL-BET combines a relational semantics for the concepts of knowledge and belief and a neighbourhood semantics for the concept of evidence. A sound and complete axiomatization for the logic is given in Sect. 3. The completeness proof is non-standard, given the interrelation between the concepts of belief and knowledge, on the one hand, and the concepts of trust and evidence, on the other hand. In Sect. 4 we discuss related work. Finally, in Sect. 5 we conclude.

2 Dynamic Logic of Belief, Evidence and Trust

In this section, we present the syntax and the semantics of the logic DL-BET and illustrate it on a concrete example.

The static component of DL-BET is called L-BET, which includes modal operators for beliefs, knowledge, trust and evidence sources *plus* special atomic formulas that allow us to represent an agent's disposition to form beliefs based on evidence, namely, how much evidence she needs to collect in support of a fact in order to have a sufficient reason to believe that the fact is true. DL-BET extends L-BET with four kinds of dynamic operators describing, respectively, (i) the consequences of an agent's public announcement, (ii) the consequences of an agent's mental operation of losing trust, (iii) the consequences of an agent's mental operation of relying on someone's judgment, and (iv) the consequences of an agent's mental operation of assessing whether a certain fact is true or false.

On the technical side, DL-BET combine methods from Dynamic Epistemic Logic (DEL) that has been developed in the past decades (cf. [5,9,16]), with those techniques from neighbourhood semantics for modal logic (cf. [12]).

2.1 Syntax

Assume a non-empty countable set of atomic propositions $Atm = \{p, q, \dots\}$ and a non-empty finite set of agent names $Agt = \{i_1, \dots, i_n\}$. Elements of 2^{Agt} are called groups (or coalitions) and are denoted by $J, J', \dots$ For every $J \in 2^{Agt}$, $|J|$ denotes the cardinality of J.

The language of DL-BET, denoted by $\mathcal{L}_{\mathsf{DL-BET}}$, is defined by the following grammar in Backus-Naur Form (BNF):

$$\alpha ::= !_i\varphi \mid -_{i,j}\varphi \mid +_{i,j}\varphi \mid ?_i\varphi$$
$$\varphi ::= p \mid \mathsf{type}(i,x,y) \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{K}_i\varphi \mid \mathsf{B}_i\varphi \mid \mathsf{E}_{i,j}\varphi \mid \mathsf{T}_{i,j}\varphi \mid [\alpha]\varphi$$

where p ranges over Atm, i, j range over Agt, x ranges over $Evd = \{k \in \mathbb{N} : 0 \leq k \leq card(Agt)\}$ and y ranges over a finite chain $Qt \subseteq [\frac{1}{2}, 1]$ such that $1 \in Qt$. Sets Evd and Qt are, respectively, the set of possible numbers of collected evidence and the set of *quota* values. Evd and Qt are finite because an agent can have a number of different evidence in support of a given fact which is at most equal to the size of Agt.

The other Boolean constructions $\top$, $\bot$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined from p, $\neg$ and $\wedge$ in the standard way.

The language of L-BET (Logic of Belief, Evidence and Trust), denoted by $\mathcal{L}_{\mathsf{L-BET}}$, is defined by:

$$\varphi ::= p \mid \mathsf{type}(i,x,y) \mid \neg\varphi \mid \varphi \wedge \psi \mid \mathsf{K}_i\varphi \mid \mathsf{B}_i\varphi \mid \mathsf{E}_{i,j}\varphi \mid \mathsf{T}_{i,j}\varphi$$

K_i is the standard modal operator of knowledge and $\mathsf{K}_i\varphi$ has to be read "agent i knows that φ is true". $\mathsf{B}_i\varphi$ has to be read "agent i believes that φ is true". The dual of the knowledge operator and the dual of the belief operator are defined as follows: $\widehat{\mathsf{K}}_i\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{K}_i\neg\varphi$ and $\widehat{\mathsf{B}}_i\varphi \stackrel{\mathtt{def}}{=} \neg\mathsf{B}_i\neg\varphi$.

$\mathsf{E}_{i,j}\varphi$ has to be read "agent i has evidence in support of φ based on the information provided by agent j".

$\mathsf{T}_{i,j}\varphi$ has to be read "agent i trusts agent j's judgement on φ". Note that, when $i = j$, the operator $\mathsf{T}_{i,j}\varphi$ captures a notion of self-trust (or self-confidence). As we mentioned earlier, since [27] similar modal operators for trust have been studied by [14,31,33]. In this paper, following [27], we use a neighbourhood semantics for interpreting the trust operators $\mathsf{T}_{i,j}$ 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 and i trusts agent j's judgement about $\neg\varphi$, without inferring that i trusts agent j's judgement about $\bot$, that is, we want formula $\mathsf{T}_{i,j}\varphi \wedge \mathsf{T}_{i,j}\neg\varphi \wedge \neg\mathsf{T}_{i,j}\bot$ to be satisfiable. This means i has potential to access j's information some of which may support φ, some of which may reject φ. For example, Bill may trust Mary's judgement about the fact that a certain stock will go upward (i.e., $\mathsf{T}_{Bill,Mary} stockUp$) and, at the same time, trust Mary's judgement about the fact that the stock will not go upward (i.e., $\mathsf{T}_{Bill,Mary}\neg stockUp$), without trusting Mary's judgement about $\bot$ (i.e., $\mathsf{T}_{Bill,Mary}\bot$).[1]

$\mathsf{type}(i,x,y)$ is a constant which characterizes agent i's epistemic type. Specifically, $\mathsf{type}(i,x,y)$ has to be read as "agent i has a level of epistemic cautiousness equal to x and an acceptance *quota* equal to y". Agent i's acceptance *quota* corresponds to the *ratio* of evidence in support of a given fact to the total amount of evidence in support of either the fact or its negation, that is required for the agent to form the belief that the fact is true. A similar notion of *quota* is studied in the area of judgment aggregation [15]. Agent i's level of epistemic cautiousness corresponds to the *amount of evidence* in support of a given fact that agent i needs to collect before forming the belief that the fact is true. As we will show below, an agent's epistemic type characterizes the agent's disposition to change her beliefs on the basis of the evidence she collects.

We distinguish four types of events: $!_i\varphi$, $-_{i,j}\varphi$, $+_{i,j}\varphi$ and $?_i\varphi$. The symbol $!_i\varphi$ denotes the event of agent i publicly announcing that φ is true. $-_{i,j}\varphi$ denotes agent i's mental operation of losing trust in agent j about φ, while the symbol $+_{i,j}\varphi$ denotes agent i's mental operation of relying on agent j's judgment about φ. We assume that if an agent loses trust in someone or relies on someone, then this fact is public (i.e., it is common knowledge that the agent has lost trust in someone/has relied on someone's judgment). Note that our logic clearly distinguishes the concept of "trusting someone's judgment", denoted by formula $\mathsf{T}_{i,j}\varphi$, from the concept of "relying on someone's judgment", denoted by events $+_{i,j}\varphi$. The former is conceived as an agent's mental attitude, while the latter is conceived as an agent's mental operation affecting her mental attitudes. Finally, $?_i\varphi$ denotes agent i's mental operation assessing whether φ is true or false. As we will show in Sect. 2.2, the latter mental operation has different possible outcomes, depending on agent i's epistemic state. In particular, agent i will be prone to *expand* her set of beliefs by φ, if she does not believe the contrary and she has sufficient reason to believe φ. She will *revise* her set of beliefs by φ, if she has sufficient reason to believe φ and currently believes the contrary.

The formula $[\alpha]\varphi$ has to be read "φ will hold after the event α takes place".

[1] As we will show in Sect. 3, formula $\neg\mathsf{T}_{i,j}\bot$ is valid in the logic DL-BET. Thus, if $\mathsf{T}_{i,j}$ was a normal modal operator, $\neg(\mathsf{T}_{i,j}\varphi \wedge \mathsf{T}_{i,j}\neg\varphi)$ would have been valid, which is highly counter-intuitive.

Let us define the following abbreviations for every $i \in Agt$ and $x \in Evd$:

$$\mathsf{E}_i^{\geq x}\varphi \stackrel{\mathtt{def}}{=} \bigvee_{J \in 2^{Agt}:|J|=x} \ \bigwedge_{j \in J} \mathsf{E}_{i,j}\varphi$$

$$\mathsf{E}_i^{x}\varphi \stackrel{\mathtt{def}}{=} \mathsf{E}_i^{\geq x}\varphi \wedge \neg \mathsf{E}_i^{\geq x+1}\varphi$$

$$\mathsf{R}_i\varphi \stackrel{\mathtt{def}}{=} \bigvee_{x,x',x'' \in Evd, y \in Qt: x > x', \frac{x}{x+x'} \geq y \text{ and } x \geq x''} \Big(\mathsf{E}_i^{x}\varphi \wedge \mathsf{E}_i^{x'}\neg\varphi \wedge \mathsf{type}(i, x'', y)\Big)$$

We use the conventions $\mathsf{E}_i^{\geq 0}\varphi \stackrel{\mathtt{def}}{=} \top$ and $\mathsf{E}_i^{\geq |Agt|+1}\varphi \stackrel{\mathtt{def}}{=} \bot$.

$\mathsf{E}_i^{\geq x}\varphi$ has to be read "agent i has at least x pieces of evidence in support of φ", whereas $\mathsf{E}_i^{x}\varphi$ has to be read "agent i has exactly x pieces of evidence in support of φ".

$\mathsf{R}_i\varphi$ has to be read "agent i has a *sufficient* reason to believe that φ is true". According to our definition, an agent has a sufficient reason to believe that φ is true if and only if:

(i) she has more evidence in support of φ than evidence in support of $\neg\varphi$,
(ii) the ratio of evidence in support of φ to the total amount of evidence in support of either φ or $\neg\varphi$, is equal to or above her acceptance quota,[2]
(iii) the amount of evidence in support of φ is equal to or above her threshold of epistemic cautiousness.

As we will highlight in Sect. 2.2, a sufficient reason to believe that φ is true ensures that the mental operation of assessing whether φ is true will result in either the expansion or the revision of the agent's set of beliefs by formula φ.

2.2 Semantics

The main notion in the semantics is given by the following definition of evidence source model which provides the basic components for the interpretation of the logic DL-BET:

Definition 1 (Evidence Source Model). *An evidence source model (ESM) is a tuple* $M = (W, E, D, S, C, T, V)$ *where:*

- W *is a set of worlds or situations;*
- $E : Agt \longrightarrow 2^{W \times W}$ *s.t. for all* $i \in Agt$, $E(i)$ *is an epistemic relation on* W*;*
- $D : Agt \longrightarrow 2^{W \times W}$ *s.t. for all* $i \in Agt$, $D(i)$ *is a doxastic relation on* W*;*
- $S : Agt \times Agt \times W \longrightarrow 2^{2^W}$ *is an evidence source function;*
- $C : Agt \times Agt \times W \longrightarrow 2^{2^W}$ *is a confidence function;*
- $T : Agt \times W \longrightarrow Evd \times Qt$ *is an epistemic type function;*
- $V : W \longrightarrow 2^{Atm}$ *is a valuation function;*

[2] Note that this ratio can be conceived as the probability that φ is true, computed of the basis of the number of evidence supporting φ.

and which satisfies the following conditions for all $i, j \in Agt$, for all $w, v \in W$ and for all $X \subseteq W$:

(C1) *$E(i)$ is an equivalence relation;*
(C2) *$D(i)$ is a serial relation;*
(C3) *$D(i) \subseteq E(i)$;*
(C4) *if $wE(i)v$ then $D(i)(w) = D(i)(v)$;*
(C5) *if $wE(i)v$ then $S(i, j, w) = S(i, j, v)$;*
(C6) *if $wE(i)v$ then $C(i, j, w) = C(i, j, v)$;*
(C7) *if $X \in C(i, j, v)$ then $X \subseteq E(i)(w)$;*
(C8) *$\emptyset \notin C(i, j, v)$;*
(C9) *if $X \in S(i, j, v)$ then $X \in C(i, j, v)$;*
(C10) *if $wE(i)v$ then $T(i, w) = T(i, v)$;*

where, for any binary relation R on W, $R(w) = \{v \in W : wRv\}$.

For notational convenience, we write E_i instead of $E(i)$ and D_i instead of $D(i)$. For every $w \in W$, $E_i(w)$ and $D_i(w)$ are called, respectively, agent i's information set and belief set at w. Agent i's information set at w is the set of worlds that agent i envisages at world w, while agent i's belief set at w is the set of worlds that agent i thinks to be possible at world w.

Constraint C1 ensures that the epistemic relation $E(i)$ is nothing but the indistinguishability relation traditionally used to model a fully introspective and truthful notion of knowledge. Constraint C2 guarantees that an agent always considers possible at least one world. This guarantees consistency of beliefs.

Constraint C3 ensures that the set of possible worlds is included in the set of envisaged worlds. Indeed, following [26], a ESM requires that an agent is capable of assessing whether an envisaged situation is *possible* or not.[3]

Constraint C4 just means that if two worlds are in the same information set of agent i, then agent i has the same belief set at these two worlds. In other words, an agent knows her beliefs.

It is worth noting that Constraints C1, C2, C3 and C4 together imply that every relation D_i is transitive and Euclidean.

$S(i, j, w)$ is the set of evidence that agent j has provided to agent i where, following [10], a piece of evidence is identified with a set of worlds. Constraint C5 means that if two worlds are in the same information set of agent i, then agent i has the same evidence at these two worlds. In other words, an agent knows her evidence.

The confidence function C specifies an agent's trust in the judgments of other agents. In particular, since each set of possible worlds X is the semantic counterpart of a formula, the meaning of $X \in C(i, j, w)$ is that, at world w, agent i trusts agent j's judgment on the truth of the formula corresponding to

[3] Here we take the term "envisaged" to be synonymous of the term "imagined". Clearly, there are situations that one can imagine that she considers impossible. For example, a person can imagine a situation in which she is the president of French republic and, at the same time, considers this situation impossible.

X. Constraint C6 means that if two worlds are in the same information set of agent i, then agent i has the same trust at these two worlds. This corresponds to a property of positive introspection for trust, i.e., an agent knows whether she trusts someone.

Constraint C7 captures compatibility between knowledge and trust. Specifically, according to Constraint C7, an agent can trust someone only about facts which are compatible with her current information set. According to Constraint C8, an agent cannot trust someone about inconsistent facts.

Constraint C9 captures the basic relationship between evidence and trust: an agent i cannot receive a piece of evidence from another agent j, unless agent i trusts agent j's judgement. In other words, trust in the source is a necessary condition for making the information provided by the source a piece of evidence. It is worth noting that Constraints C7 and C9 together imply that an agent can have evidence only about facts which are compatible with her current information set, while Constraints C8 and C9 together imply that an agent cannot have evidence about inconsistent facts, that is:

- if $X \in S(i,j,v)$ then $X \subseteq E(i)(w)$, and
- $\emptyset \notin S(i,j,v)$.

$T(i,w)$ corresponds to agent i's epistemic type at world w. Constraint C10 just means that if two worlds are in the same information set of agent i, then agent i has the same epistemic type at these two worlds. In other words, an agent knows her epistemic type. As emphasized above, an agent's epistemic type is defined by the agent's level of epistemic cautiousness and the agent's acceptance *quota*.

Truth conditions of DL-BET formulas are inductively defined as follows.

Definition 2 (Truth conditions). *Let $M = (W, E, D, S, C, T, V)$ be a ESM and let $w \in W$. Then:*

$$\begin{aligned}
M, w \models p &\Longleftrightarrow p \in V(w)\\
M, w \models \mathsf{type}(i,x,y) &\Longleftrightarrow T(i,w) = (x,y)\\
M, w \models \neg\varphi &\Longleftrightarrow M, w \not\models \varphi\\
M, w \models \varphi \wedge \psi &\Longleftrightarrow M, w \models \varphi \text{ and } M, w \models \psi\\
M, w \models \mathsf{K}_i\varphi &\Longleftrightarrow \forall v \in E_i(w) : M, v \models \varphi\\
M, w \models \mathsf{B}_i\varphi &\Longleftrightarrow \forall v \in D_i(w) : M, v \models \varphi\\
M, w \models \mathsf{E}_{i,j}\varphi &\Longleftrightarrow ||\varphi||^M_{i,w} \in S(i,j,w)\\
M, w \models \mathsf{T}_{i,j}\varphi &\Longleftrightarrow ||\varphi||^M_{i,w} \in C(i,j,w)\\
M, w \models [\alpha]\psi &\Longleftrightarrow M^{\alpha}, w \models \psi
\end{aligned}$$

where

$$||\varphi||^M_{i,w} = \{v \in W : M, v \models \varphi\} \cap E_i(w),$$

$M^{!_i\varphi}$, $M^{-_{i,j}\varphi}$, $M^{+_{i,j}\varphi}$ and $M^{?_i\varphi}$ are updated models defined according to the following Definitions 3, 4, 5 and 6.

According to the truth conditions: agent i knows that φ at world w if and only if φ is true in all worlds that at w agent i envisages, and agent i believes that φ at world w if and only if φ is true in all worlds that at w agent i considers possible. Moreover, at world w agent j has provided evidence in support of φ to agent i if and only if, at w, agent i has the fact corresponding to the formula φ (i.e., $||\varphi||^M_{i,w}$) included in her evidence set $S(i, j, w)$. Finally, at world w agent i trusts agent j's judgment about φ if and only if, at w, the fact corresponding to the formula φ (i.e., $||\varphi||^M_{i,w}$) is included in agent i's confidence set $C(i, j, w)$. In what follows, we define the updated models triggered by the four kinds of events:

Definition 3 (Update via $!_i\varphi$). *Let $M = (W, E, D, S, C, T, V)$ be a ESM. Then, $M^{!_i\varphi}$ is the tuple $(W, E, D, S^{!_i\varphi}, C, T, V)$ such that, for all $j, k \in Agt$ and $w \in W$:*

$$S^{!_i\varphi}(j,k,w) = \begin{cases} S(j,k,w) \cup \{||\varphi||^M_{j,w}\} & \text{if } k = i \text{ and } M, w \models \mathsf{T}_{j,i}\varphi \\ S(j,k,w) & \text{otherwise} \end{cases}$$

Definition 4 (Update via $-_{i,j}\varphi$). *Let $M = (W, E, D, S, C, T, V)$ be a ESM. Then, $M^{-_{i,j}\varphi}$ is the tuple $(W, E, D, S^{-_{i,j}\varphi}, C^{-_{i,j}\varphi}, T, V)$ such that, for all $k, l \in Agt$ and $w \in W$:*

$$S^{-_{i,j}\varphi}(k,l,w) = \begin{cases} S(k,l,w) \setminus \{||\varphi||^M_{k,w}\} & \text{if } k = i \text{ and } l = j \\ S(k,l,w) & \text{otherwise} \end{cases}$$

$$C^{-_{i,j}\varphi}(k,l,w) = \begin{cases} C(k,l,w) \setminus \{||\varphi||^M_{k,w}\} & \text{if } k = i \text{ and } l = j \\ C(k,l,w) & \text{otherwise} \end{cases}$$

Definition 5 (Update via $+_{i,j}\varphi$). *Let $M = (W, E, D, S, C, T, V)$ be a ESM. Then, $M^{+_{i,j}\varphi}$ is the tuple $(W, E, D, S, C^{+_{i,j}\varphi}, T, V)$ such that, for all $k, l \in Agt$ and $w \in W$:*

$$C^{+_{i,j}\varphi}(k,l,w) = \begin{cases} C(k,l,w) \cup \{||\varphi||^M_{k,w}\} & \text{if } k = i \text{ and } l = j \text{ and } M, w \models \widehat{\mathsf{K}}_i\varphi \\ C(k,l,w) & \text{otherwise} \end{cases}$$

Definition 6 (Update via $?_i\varphi$). *Let $M = (W, E, D, S, C, T, V)$ be a ESM. Then, $M^{?_i\varphi}$ is the tuple $(W, E, D^{?_i\varphi}, S, C, T, V)$ such that, for all $j \in Agt$ and $w \in W$:*

$$D_j^{?_i\varphi}(w) = \begin{cases} D_j(w) \cap ||\varphi||_{j,w}^M & \textit{if } j = i \textit{ and } M, w \models \mathsf{R}_i\varphi \wedge \neg\mathsf{B}_i\neg\varphi \\ D_j(w) \cap ||\neg\varphi||_{j,w}^M & \textit{if } j = i \textit{ and } M, w \models \mathsf{R}_i\neg\varphi \wedge \neg\mathsf{B}_i\varphi \\ ||\varphi||_{j,w}^M & \textit{if } j = i \textit{ and } M, w \models \mathsf{R}_i\varphi \wedge \mathsf{B}_i\neg\varphi \\ ||\neg\varphi||_{j,w}^M & \textit{if } j = i \textit{ and } M, w \models \mathsf{R}_i\neg\varphi \wedge \mathsf{B}_i\varphi \\ D_j(w) & \textit{otherwise} \end{cases}$$

As highlighted by Definition 3, if an agent announces that φ is true, then she will provide a piece of new evidence in support of φ only to the agents who trust her judgement about φ.

According to Definition 4, if an agent i loses trust in someone about a given fact, then this fact is removed from the set of facts for which i trusts j's judgment. To ensure that Constraint C9 in Definition 1 is preserved under this model update operation, the fact is also removed from the set of evidence provided by agent j to agent i.

Acccording to Definition 5, if agent i relies on agent j's judgment about a certain fact then, as a consequence, this fact is added to the set of facts for which i trusts j's judgment, under the condition that the fact is consistent with i's knowledge. The latter condition guarantees that Constraint C8 in Definition 1 is preserved under this model update operation.

According to Definition 6, the mental operation of assessing whether φ is true has five possible outcomes:

- if an agent has a sufficient reason to believe that φ is true and does not believe that φ is false, then she *expands* her beliefs by removing from her belief set the worlds in which φ is false,
- if an agent has a sufficient reason to believe that φ is false and does not believe that φ is true, then she *expands* her beliefs by removing from her belief set the worlds in which φ is true,
- if an agent has a sufficient reason to believe that φ is true and actually believes that φ is false, then she *revises* her beliefs by removing from her belief set the worlds in which φ is false and including all worlds of her information set in which φ is true,
- if an agent has a sufficient reason to believe that φ is false and actually believes that φ is true, then she *revises* her beliefs by removing from her belief set the worlds in which φ is true and including all worlds of her information set in which φ is false,
- if an agent has no sufficient reason to believe that φ is true and has no sufficient reason to believe that φ is false, then she *suspends her judgement* about φ and does not change her belief set.

This highlights the distinction between *expansion*, *revision* and *suspension of judgement*. Since [1], the former two mental operations have been extensively studied in the area of belief revision. While expansion captures the idea of increasing the set of facts that an agent believes, revision captures the idea of restoring consistency, after having added to the set of beliefs a new information that is inconsistent with the pre-existing information. The latter mental operation has been studied in the epistemological area (see, e.g., [20]). It captures the idea that an agent is not willing to integrate a new information in her set of beliefs, unless she has gathered enough evidence in support of it.

As the following proposition highlights, our model update operations are well-defined as they preserve the properties of evidence source models (ESMs) as defined in Definition 1.

Proposition 1. *If M is a ESM then $M^{!_i\varphi}$, $M^{-_{i,j}\varphi}$, $M^{+_{i,j}\varphi}$ and $M^{?_i\varphi}$ are ESMs too.*

For every $\varphi \in \mathcal{L}_{\mathsf{DL-BET}}$, we write $\models \varphi$ to mean that φ is valid w.r.t. the class of ESMs, that is, for every $M = (W, E, D, S, C, T, V)$ and for every $w \in W$ we have $M, w \models \varphi$. We say that φ is satisfiable w.r.t. the class of ESMs if and only if $\neg\varphi$ is not valid w.r.t. the class of ESMs.

2.3 Some Properties

In this section we focus on some basic properties of the logic DL-BET. We start with the following static properties of evidence, trust and reason:

$$\models \mathsf{T}_{i,j}\varphi \rightarrow \widehat{\mathsf{K}}_i\varphi \tag{1}$$

$$\models \mathsf{E}_{i,j}\varphi \rightarrow \widehat{\mathsf{K}}_i\varphi \tag{2}$$

$$\models \mathsf{R}_i\varphi \rightarrow \widehat{\mathsf{K}}_i\varphi \tag{3}$$

$$\models \neg(\mathsf{R}_i\varphi \wedge \mathsf{R}_i\neg\varphi) \tag{4}$$

According to the validities (1), (2) and (3), trust, evidence and reason are always consistent with knowledge. The validity (4) highlights that an agent cannot have inconsistent reasons.

Let us now consider some dynamic properties that only apply to the propositional fragment of the logic DL-BET. Let $\mathcal{L}_{Atm}$ be the propositional language build out of the set of atoms Atm. Then, for $\varphi, \psi \in \mathcal{L}_{Atm}$ we have:

$$\models \mathsf{T}_{i,j}\varphi \rightarrow [!_j\varphi]\mathsf{E}_{i,j}\varphi \tag{5}$$

$$\models [-_{i,j}\varphi](\neg\mathsf{E}_{i,j}\varphi \wedge \neg\mathsf{T}_{i,j}\varphi) \tag{6}$$

$$\models \widehat{\mathsf{K}}_i\varphi \rightarrow [+_{i,j}\varphi]\mathsf{T}_{i,j}\varphi \tag{7}$$

$$\models \mathsf{R}_i\varphi \rightarrow [?_i\varphi]\mathsf{B}_i\varphi \tag{8}$$

$$\models \mathsf{R}_i\neg\varphi \rightarrow [?_i\varphi]\mathsf{B}_i\neg\varphi \tag{9}$$

$$\models \big((\mathsf{R}_i\varphi \wedge \mathsf{K}_i(\varphi \rightarrow \psi)) \vee (\mathsf{R}_i\neg\varphi \wedge \mathsf{K}_i(\neg\varphi \rightarrow \psi))\big) \rightarrow [?_i\varphi]\mathsf{B}_i\psi \tag{10}$$

$$\models \big((\mathsf{R}_i\varphi \wedge \neg\mathsf{B}_i\neg\varphi \wedge \mathsf{B}_i\psi) \vee (\mathsf{R}_i\neg\varphi \wedge \neg\mathsf{B}_i\varphi \wedge \mathsf{B}_i\psi)\big) \rightarrow [?_i\varphi]\mathsf{B}_i\psi \tag{11}$$

According to the validity (5), if an agent trusts the information source's judgment about φ, then she will have an additional evidence in support of φ after the information source has publicly announced that φ is true. Validities (6) and (7) highlight the basic properties of the mental operation of losing trust in someone's judgment and relying on someone's judgment. Specifically, after having lost trust in agent j's judgment about φ, i does not trust anymore j's judgment about φ and j cannot provide any more evidence in support of φ. Moreover, if φ is consistent with agent i's knowledge then, after having relied on agent j's judgment about φ, i does trusts j's judgment about φ. Validities (8) and (9) highlight the role of reason in the formation of belief: if agent i has a sufficient reason to believe $\varphi/\neg\varphi$ then, after having assessed whether φ is true, she will start to believe $\varphi/\neg\varphi$. Validity (10) highlights the role of knowledge in reason-based belief change: if an agent has a sufficient reason to believe $\varphi/\neg\varphi$ and knows that $\varphi/\neg\varphi$ implies ψ then, after having assessed whether φ is true, she will start to believe ψ. Validity (11) highlights the conservative aspect of reason-based belief expansion: if an agent has a sufficient reason to believe $\varphi/\neg\varphi$ without believing the contrary and believes ψ then, after having assessed whether φ is true, she will continue to believe ψ.

The reason why we need to impose that φ and ψ are propositional formulas is that there are DL-BET-formulas such as the Moore-like formula $p \wedge \neg \mathsf{B}_i p$ for which the previous validities (5)–(11) do not hold. For instance, the following formula is not valid:

$$\mathsf{R}_i(p \wedge \neg \mathsf{B}_i p) \rightarrow [?_i(p \wedge \neg \mathsf{B}_i p)]\mathsf{B}_i(p \wedge \neg \mathsf{B}_i p).$$

This is intuitive since if I have sufficient reason to believe that my uncertainty about p could be unjustified then, after assessing whether this is the case, I may start to believe that p and that I believe this (since I have introspection over my beliefs).

2.4 An Example

This section is devoted to illustrate the syntax and the semantics of the logic DL-BET with the aid of a concrete example of AI application.

Suppose a human user wants to know whether the movie The Tree of Life by Terrence Malick is a great movie or not and asks this to her chatbot. The chabot has access to four information sources in the Internet, namely, Wikipedia, IMDb, Amazon and Rotten Tomatoes (RT). The chatbot knows that if "The Tree of Life has won the Palm d'Or at the Cannes festival", denoted by proposition p, then "The Tree of Life is a great movie", denoted by proposition q:

$$Hyp1 \stackrel{\text{def}}{=} \mathsf{K}_{chatbot}(p \rightarrow q)$$

Moreover, the chatbot trusts the judgments of both RT and Amazon about q and $\neg q$. This means that if either RT or Amazon says that The Tree of Life is a great movie/is not a great movive, then this counts as a piece of evidence in support of this fact. Finally, the chatbot trusts the judgments of both Wikipedia and IMDb about p and $\neg p$:

$$\begin{aligned} \mathit{Hyp2} \overset{\text{def}}{=}\ & \mathsf{T}_{chatbot, Wikipedia}p \wedge \mathsf{T}_{chatbot,IMDb}p \wedge \\ & \mathsf{T}_{chatbot, Wikipedia}\neg p \wedge \mathsf{T}_{chatbot,IMDb}\neg p \wedge \\ & \mathsf{T}_{chatbot,RT}q \wedge \mathsf{T}_{chatbot,Amazon}q \wedge \\ & \mathsf{T}_{chatbot,RT}\neg q \wedge \mathsf{T}_{chatbot,Amazon}\neg q \end{aligned}$$

Moreover, suppose that the chatbot (i) is uncertain whether p is true and is uncertain whether q is true, (ii) has no evidence in support of $p, \neg p, q$ and $\neg q$, and (iii) has a level of epistemic cautiousness equal to 2 and an acceptance *quota* equal to 1. That is:

$$\begin{aligned} \mathit{Hyp3} \overset{\text{def}}{=}\ & \neg\mathsf{B}_{chatbot}p \wedge \neg\mathsf{B}_{chatbot}\neg p \wedge \neg\mathsf{B}_{chatbot}q \wedge \neg\mathsf{B}_{chatbot}\neg q \wedge \\ & \mathsf{E}^0_{chatbot}p \wedge \mathsf{E}^0_{chatbot}\neg p \wedge \mathsf{E}^0_{chatbot}q \wedge \mathsf{E}^0_{chatbot}\neg q \wedge \mathsf{type}(chatbot, 2, 1) \end{aligned}$$

Thus, if the chatbot learns from Amazon that The Tree of Life is a great movie while it learns from RT that The Tree of Life is not a great movie, it will be unable to draw any conclusion about the fact that The Tree of Life is a great movie and keep its initial uncertainty. That is:

$$\models (\mathit{Hyp1} \wedge \mathit{Hyp2} \wedge \mathit{Hyp3}) \rightarrow\!| \ [!_{Amazon}q][!_{RT}\neg q][?_{chatbot}q](\neg\mathsf{B}_{chatbot}q \wedge \neg\mathsf{B}_{chatbot}\neg q)$$

On the contrary, if the chatbot learns both from Wikipedia and from IMDb that The Tree of Life has won the Palm d'Or at the Cannes festival, then it will be able to infer that The Tree of Life is a great movie. That is:

$$\models (\mathit{Hyp1} \wedge \mathit{Hyp2} \wedge \mathit{Hyp3}) \rightarrow [!_{Wikipedia}p][!_{IMDb}p][?_{chatbot}p]\mathsf{B}_{chatbot}q$$

3 Axiomatization

Let us now present sound and complete axiomatizations for the logic L-BET and its dynamic extension DL-BET. The completeness proof of L-BET is based on a canonical model construction.[4] All axioms of L-BET, except two, are used in the usual way to prove that the canonical model so constructed is a ESM. There are two special axioms of the logic L-BET, about the interrelation between knowledge and trust and between knowledge and evidence that are used in an unusual way to prove the truth lemma.

Definition 7 (L-BET). *We define L-BET to be the extension of classical propositional logic given by the following rules and axioms:*

[4] The proof can be found in the extended version of this paper [28].

$$(\mathsf{K}_i\varphi \wedge \mathsf{K}_i(\varphi \rightarrow \psi)) \rightarrow \mathsf{K}_i\psi \qquad (\mathbf{K}_{\mathsf{K}_i})$$
$$\mathsf{K}_i\varphi \rightarrow \varphi \qquad (\mathbf{T}_{\mathsf{K}_i})$$
$$\mathsf{K}_i\varphi \rightarrow \mathsf{K}_i\mathsf{K}_i\varphi \qquad (\mathbf{4}_{\mathsf{K}_i})$$
$$\neg\mathsf{K}_i\varphi \rightarrow \mathsf{K}_i\neg\mathsf{K}_i\varphi \qquad (\mathbf{5}_{\mathsf{K}_i})$$
$$(\mathsf{B}_i\varphi \wedge \mathsf{B}_i(\varphi \rightarrow \psi)) \rightarrow \mathsf{B}_i\psi \qquad (\mathbf{K}_{\mathsf{B}_i})$$
$$\neg(\mathsf{B}_i\varphi \wedge \mathsf{B}_i\neg\varphi) \qquad (\mathbf{D}_{\mathsf{B}_i})$$
$$\neg\mathsf{T}_{i,j}\bot \qquad (\mathbf{Cons}_{\mathsf{T}_{i,j}})$$
$$\bigvee_{x\in Evd, y\in Qt} \mathsf{type}(i,x,y) \qquad (\mathbf{AtLeast}_{\mathsf{type}(i,x,y)})$$
$$\mathsf{type}(i,x,y) \rightarrow \neg\mathsf{type}(i,x',y') \textit{ if } x \neq x' \textit{ or } y \neq y' \qquad (\mathbf{AtMost}_{\mathsf{type}(i,x,y)})$$
$$\mathsf{K}_i\varphi \rightarrow \mathsf{B}_i\varphi \qquad (\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$$
$$\mathsf{B}_i\varphi \rightarrow \mathsf{K}_i\mathsf{B}_i\varphi \qquad (\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$$
$$\mathsf{type}(i,x,y) \rightarrow \mathsf{K}_i\mathsf{type}(i,x,y) \qquad (\mathbf{Mix}_{\mathsf{K}_i,\mathsf{type}(i,x,y)})$$
$$\mathsf{E}_{i,j}\varphi \rightarrow \mathsf{T}_{i,j}\varphi \qquad (\mathbf{Mix}_{\mathsf{E}_{i,j},\mathsf{T}_{i,j}})$$
$$\mathsf{E}_{i,j}\varphi \rightarrow \mathsf{K}_i\mathsf{E}_{i,j}\varphi \qquad (\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{E}_{i,j}})$$
$$\mathsf{T}_{i,j}\varphi \rightarrow \mathsf{K}_i\mathsf{T}_{i,j}\varphi \qquad (\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{T}_{i,j}})$$
$$(\mathsf{T}_{i,j}\varphi \wedge \mathsf{K}_i(\varphi \leftrightarrow \psi)) \rightarrow \mathsf{T}_{i,j}\psi \qquad (\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{T}_{i,j}})$$
$$(\mathsf{E}_{i,j}\varphi \wedge \mathsf{K}_i(\varphi \leftrightarrow \psi)) \rightarrow \mathsf{E}_{i,j}\psi \qquad (\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{E}_{i,j}})$$
$$\frac{\varphi}{\mathsf{K}_i\varphi} \qquad (\mathbf{Nec}_{\mathsf{K}_i})$$

Note that the rule of necessitation for B_i is provable by $(\mathbf{Nec}_{\mathsf{K}_i})$ and $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$. Moreover, Axiom 4 for B_i is provable by $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$ and $(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$. Axiom 5 for B_i is provable by means of $(\mathbf{Mix1}_{\mathsf{K}_i,\mathsf{B}_i})$, $(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{B}_i})$, $\mathbf{K}_{\mathsf{K}_i}$, $\mathbf{T}_{\mathsf{K}_i}$, $\mathbf{4}_{\mathsf{K}_i}$ and $\mathbf{5}_{\mathsf{K}_i}$. A syntactic proof can be found in [30]. Finally, the following rules of equivalence for trust and evidence are provable by means of $(\mathbf{Nec}_{\mathsf{K}_i})$, $(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{T}_{i,j}})$ and $(\mathbf{Mix2}_{\mathsf{K}_i,\mathsf{E}_{i,j}})$:

$$\frac{\varphi \leftrightarrow \psi}{\mathsf{T}_{i,j}\varphi \leftrightarrow \mathsf{T}_{i,j}\psi} \qquad (12)$$

$$\frac{\varphi \leftrightarrow \psi}{\mathsf{E}_{i,j}\varphi \leftrightarrow \mathsf{E}_{i,j}\psi} \qquad (13)$$

Theorem 1. *The logic* L-BET *is sound and complete for the class of ESMs.*

The axiomatics of the logic DL-BET includes all principles of the logic L-BET *plus* a set of reduction axioms and the rule of replacement of equivalents.

Definition 8. *We define DL-BET to be the extension of L-BET generated by the following reduction axioms for the dynamic operators* $[!_i\varphi]$*:*

$$[!_i\varphi]p \leftrightarrow p \quad (\mathbf{Red}_{!_i\varphi,p})$$
$$[!_i\varphi]\mathsf{type}(k,x,y) \leftrightarrow \mathsf{type}(k,x,y) \quad (\mathbf{Red}_{!_i\varphi,\mathsf{type}(l,x,y)})$$
$$[!_i\varphi]\neg\psi \leftrightarrow \neg[!_i\varphi]\psi \quad (\mathbf{Red}_{!_i\varphi,\neg})$$
$$[!_i\varphi](\psi \wedge \chi) \leftrightarrow ([!_i\varphi]\psi \wedge [!_i\varphi]\chi) \quad (\mathbf{Red}_{!_i\varphi,\wedge})$$
$$[!_i\varphi]\mathsf{K}_j\psi \leftrightarrow \mathsf{K}_j[!_i\varphi]\psi \quad (\mathbf{Red}_{!_i\varphi,\mathsf{K}_j})$$
$$[!_i\varphi]\mathsf{B}_j\psi \leftrightarrow \mathsf{B}_j[!_i\varphi]\psi \quad (\mathbf{Red}_{!_i\varphi,\mathsf{B}_j})$$
$$[!_i\varphi]\mathsf{E}_{j,k}\psi \leftrightarrow \mathsf{E}_{j,k}[!_i\varphi]\psi \text{ if } i \neq k \quad (\mathbf{Red}_{!_i\varphi,\mathsf{E}_{j,k}})$$
$$[!_i\varphi]\mathsf{E}_{j,i}\psi \leftrightarrow \Big(\big(\mathsf{T}_{j,i}\varphi \rightarrow (\mathsf{E}_{j,i}[!_i\varphi]\psi \vee \mathsf{K}_j(\varphi \leftrightarrow [!_i\varphi]\psi))\big)\wedge \big(\neg\mathsf{T}_{j,i}\varphi \rightarrow \mathsf{E}_{j,i}[!_i\varphi]\psi\big)\Big) \quad (\mathbf{Red}_{!_i\varphi,\mathsf{E}_{j,i}})$$
$$[!_i\varphi]\mathsf{T}_{j,k}\psi \leftrightarrow \mathsf{T}_{j,k}[!_i\varphi]\psi \quad (\mathbf{Red}_{!_i\varphi,\mathsf{T}_{j,k}})$$

the following ones for the dynamic operators $[-_{i,j}\varphi]$*:*

$$[-_{i,j}\varphi]p \leftrightarrow p \quad (\mathbf{Red}_{-_{i,j}\varphi,p})$$
$$[-_{i,j}\varphi]\mathsf{type}(k,x,y) \leftrightarrow \mathsf{type}(k,x,y) \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{type}(l,x,y)})$$
$$[-_{i,j}\varphi]\neg\psi \leftrightarrow \neg[-_{i,j}\varphi]\psi \quad (\mathbf{Red}_{-_{i,j}\varphi,\neg})$$
$$[-_{i,j}\varphi](\psi \wedge \chi) \leftrightarrow ([-_{i,j}\varphi]\psi \wedge [-_{i,j}\varphi]\chi) \quad (\mathbf{Red}_{-_{i,j}\varphi,\wedge})$$
$$[-_{i,j}\varphi]\mathsf{K}_k\psi \leftrightarrow \mathsf{K}_k[-_{i,j}\varphi]\psi \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{K}_k})$$
$$[-_{i,j}\varphi]\mathsf{B}_k\psi \leftrightarrow \mathsf{B}_k[-_{i,j}\varphi]\psi \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{B}_k})$$
$$[-_{i,j}\varphi]\mathsf{E}_{k,l}\psi \leftrightarrow \mathsf{E}_{k,l}[-_{i,j}\varphi]\psi \text{ if } i \neq k \text{ or } j \neq l \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{E}_{k,l}})$$
$$[-_{i,j}\varphi]\mathsf{E}_{i,j}\psi \leftrightarrow \big(\mathsf{E}_{i,j}[-_{i,j}\varphi]\psi \vee \neg\mathsf{K}_i(\varphi \leftrightarrow [-_{i,j}\varphi]\psi)\big) \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{E}_{i,j}})$$
$$[-_{i,j}\varphi]\mathsf{T}_{k,l}\psi \leftrightarrow \mathsf{T}_{k,l}[-_{i,j}\varphi]\psi \text{ if } i \neq k \text{ or } j \neq l \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{T}_{k,l}})$$
$$[-_{i,j}\varphi]\mathsf{T}_{i,j}\psi \leftrightarrow \big(\mathsf{T}_{i,j}[-_{i,j}\varphi]\psi \vee \neg\mathsf{K}_i(\varphi \leftrightarrow [-_{i,j}\varphi]\psi)\big) \quad (\mathbf{Red}_{-_{i,j}\varphi,\mathsf{T}_{i,j}})$$

the following ones for the dynamic operators $[+_{i,j}\varphi]$*:*

$$[+_{i,j}\varphi]p \leftrightarrow p \quad (\mathbf{Red}_{+_{i,j}\varphi,p})$$
$$[+_{i,j}\varphi]\mathsf{type}(k,x,y) \leftrightarrow \mathsf{type}(k,x,y) \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{type}(l,x,y)})$$
$$[+_{i,j}\varphi]\neg\psi \leftrightarrow \neg[+_{i,j}\varphi]\psi \quad (\mathbf{Red}_{+_{i,j}\varphi,\neg})$$
$$[+_{i,j}\varphi](\psi \wedge \chi) \leftrightarrow ([+_{i,j}\varphi]\psi \wedge [+_{i,j}\varphi]\chi) \quad (\mathbf{Red}_{+_{i,j}\varphi,\wedge})$$
$$[+_{i,j}\varphi]\mathsf{K}_k\psi \leftrightarrow \mathsf{K}_k[+_{i,j}\varphi]\psi \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{K}_k})$$
$$[+_{i,j}\varphi]\mathsf{B}_k\psi \leftrightarrow \mathsf{B}_k[+_{i,j}\varphi]\psi \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{B}_k})$$
$$[+_{i,j}\varphi]\mathsf{E}_{k,l}\psi \leftrightarrow \mathsf{E}_{k,l}[+_{i,j}\varphi]\psi \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{E}_{k,l}})$$
$$[+_{i,j}\varphi]\mathsf{T}_{k,l}\psi \leftrightarrow \mathsf{T}_{k,l}[+_{i,j}\varphi]\psi \text{ if } i \neq k \text{ or } j \neq l \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{T}_{k,l}})$$
$$[+_{i,j}\varphi]\mathsf{T}_{i,j}\psi \leftrightarrow \Big(\big(\widehat{\mathsf{K}}_i\varphi \rightarrow (\mathsf{T}_{i,j}[+_{i,j}\varphi]\psi \vee \mathsf{K}_i(\varphi \leftrightarrow [+_{i,j}\varphi]\psi))\big)\wedge \big(\mathsf{K}_i\neg\varphi \rightarrow \mathsf{T}_{i,j}[+_{i,j}\varphi]\psi\big)\Big) \quad (\mathbf{Red}_{+_{i,j}\varphi,\mathsf{T}_{i,j}})$$

the following ones for the dynamic operators $[?_i\varphi]$*:*

$$[?_i\varphi]p \leftrightarrow p \quad (\mathbf{Red}_{?_i\varphi,p})$$

$$[?_i\varphi]\mathsf{type}(k,x,y) \leftrightarrow \mathsf{type}(k,x,y) \quad (\mathbf{Red}_{?_i\varphi,\mathsf{type}(l,x,y)})$$

$$[?_i\varphi]\neg\psi \leftrightarrow \neg[?_i\varphi]\psi \quad (\mathbf{Red}_{?_i\varphi,\neg})$$

$$[?_i\varphi](\psi \wedge \chi) \leftrightarrow ([?_i\varphi]\psi \wedge [?_i\varphi]\chi) \quad (\mathbf{Red}_{?_i\varphi,\wedge})$$

$$[?_i\varphi]\mathsf{K}_j\psi \leftrightarrow \mathsf{K}_j[?_i\varphi]\psi \quad (\mathbf{Red}_{?_i\varphi,\mathsf{K}_j})$$

$$[?_i\varphi]\mathsf{B}_j\psi \leftrightarrow \mathsf{B}_j[?_i\varphi]\psi \text{ if } i \neq j \quad (\mathbf{Red}_{?_i\varphi,\mathsf{B}_j})$$

$$[?_i\varphi]\mathsf{B}_i\psi \leftrightarrow \Big(\big(\alpha_1 \to \mathsf{B}_i(\varphi \to [?_i\varphi]\psi)\big)\wedge \big(\alpha_2 \to \mathsf{B}_i(\neg\varphi \to [?_i\varphi]\psi)\big)\wedge \big(\alpha_3 \to \mathsf{K}_i(\varphi \to [?_i\varphi]\psi)\big)\wedge \big(\alpha_4 \to \mathsf{K}_i(\neg\varphi \to [?_i\varphi]\psi)\big)\wedge \big(\alpha_5 \to \mathsf{B}_i[?_i\varphi]\psi\big)\Big) \quad (\mathbf{Red}_{?_i\varphi,\mathsf{B}_i})$$

$$[?_i\varphi]\mathsf{E}_{j,k}\psi \leftrightarrow \mathsf{E}_{j,k}[?_i\varphi]\psi \quad (\mathbf{Red}_{?_i\varphi,\mathsf{E}_{j,k}})$$

$$[?_i\varphi]\mathsf{E}_{j,k}\psi \leftrightarrow \mathsf{E}_{j,k}[?_i\varphi]\psi \quad (\mathbf{Red}_{?_i\varphi,\mathsf{E}_{j,k}})$$

$$[?_i\varphi]\mathsf{T}_{j,k}\psi \leftrightarrow \mathsf{T}_{j,k}[?_i\varphi]\psi \quad (\mathbf{Red}_{?_i\varphi,\mathsf{T}_{j,k}})$$

and the following rule of inference:

$$\frac{\psi_1 \leftrightarrow \psi_2}{\varphi \leftrightarrow \varphi[\psi_1/\psi_2]} \quad (\mathbf{RRE})$$

with:

$$\alpha_1 \stackrel{\text{def}}{=} \mathsf{R}_i\varphi \wedge \neg\mathsf{B}_i\neg\varphi$$

$$\alpha_2 \stackrel{\text{def}}{=} \mathsf{R}_i\neg\varphi \wedge \neg\mathsf{B}_i\varphi$$

$$\alpha_3 \stackrel{\text{def}}{=} \mathsf{R}_i\varphi \wedge \mathsf{B}_i\neg\varphi$$

$$\alpha_4 \stackrel{\text{def}}{=} \mathsf{R}_i\neg\varphi \wedge \mathsf{B}_i\varphi$$

$$\alpha_5 \stackrel{\text{def}}{=} \neg\alpha_1 \wedge \neg\alpha_2 \wedge \neg\alpha_3 \wedge \neg\alpha_4$$

The completeness of DL-BET follows from Theorem 1, in view of the fact that the reduction axioms and the rule (**RRE**) may be used to find, for any DL-BET formula, a provably equivalent L-BET formula.

Theorem 2. *DL-BET is sound and complete for the class of ESMs.*

4 Related Work

Artemov [2] proposes so-called justification logic in which evidence is expressed as a term, and possible manipulations of evidence are operations over terms. This framework has been further connected to the notion of explicit and implicit beliefs and belief revision in [6]. Differently, [10,11] adopts a neighbourhood semantics, adding evidence to the standard belief model in the form of families of sets of possible worlds, and studies evidence-based belief change. These approaches share with DL-BET the idea of modeling the relationship between evidence and belief.

Social influence in terms of individual's belief change has caught a lot of attention in recent years. Liu, Seligman and Girard [29] proposes a finite state automata model with a threshold to deal with social influence. As a simple case, agent i would change her belief from p to $\neg p$ if all her neighbors believe $\neg p$. This model can successfully explain social phenomena, like peer pressure, and behavior adoption. Christoff [13] further develops this model and investigates various features of social networks and their evolution over time, including information flow and spread of opinions. Xue and Parikh [34] looks at expert influence in social network, and show how an agent makes decisions when facing conflicting choices in belief update. These approaches share with DL-BET the idea of modeling belief change and belief formation due to the information received by, possibly conflicting, information sources.

When considering the relationship between agents in the context of information exchange, trust is the core notion in play. Early work [27] studies the influence of trust on agent's formation of beliefs with an axiom saying that if agent i believes that agent j has told her the truth about p, and she trusts the judgement of j about p, then she will also believe p. In the context of social influence, [4] introduces quantitative measurement on trust between agents and strength of evidence, and stipulates how these parameters influence one's valuation of new evidence. Lorini, Jiang and Perrussel [31] studies the phenomenon of trust-based belief change, that is, belief change that depends on the degree of trust the receiver has in the source of information. In a similar way, in [22] trust is conceived as a pre-processing step before belief revision. Viewed in line of social choice theory, one can also think of belief formation or change as a process of aggregating opinions from different reliable information sources, as in [21]. These approaches share with DL-BET the idea that trust in the information source plays a fundamental role in belief change and belief formation.

In the area of information fusion, a similar concern on belief, evidence and trust has led to number of proposals [18,23,24,32]. However, these approaches lean heavily on the machinery of Bayesian probability theory. For instance, in so-called subjective logic, based on the Dempster-Shafer rule, Jøsang proposes a new Bayesian update function to study belief revision. In contrast, our main concern with DL-BET is the logical relationship between notions of knowledge, belief, evidence and trust, as well as the principles of reasoning about them, thus our work is qualitative in nature.

5 Conclusion

In this paper we have proposed a new logic, called "Dynamic Logic of Belief, Evidence and Trust" (DL-BET), which supports reasoning about evidence-based belief formation and belief change in a multi-agent setting. We have provided a complete axiomatization for both the static L-BET and its dynamic extension DL-BET. We have illustrated the expressive power of DL-BET with the aid of a concrete example involving a chatbot interacting with a human.

The logical account of belief revision we have provided (Definition 6) is radical: if an agent has a sufficient reason to believe a certain fact φ and currently believes the opposite then, after assessing whether φ is true, she will include in her belief set *all worlds* of her information set in which φ is true. In future work, we plan to extend the formal semantics of the logic DL-BET by a plausibility ordering over possible words for each agent, as traditionally used in modal logic analysis of belief revision (see, e.g., [7,8]). This extension will allow us to refine the belief revision operation by assuming that, after a revision by φ, an agent will include in her belief set only the *best worlds* (according to the plausibility ordering) of her information set in which φ is true.

We have emphasized that the epistemic cautiousness level and acceptance quota specify together how much evidence one needs to collect to form a belief, or change one's belief. Though our framework has worked with numerals already, we have managed to get a complete logic. In future work we would like to extend our model to handle uncertainties in evidence and belief as well as with degrees of trust or graded trust in information sources. It is our intention to explore how far we can go with such a still qualitative-oriented approach against a completely quantitative method.

Finally, an agent obtains information from trusted sources by social communication, and forms her beliefs on the basis of reasons. In this paper, we have investigated epistemic reasons. We plan to extend our logical framework with agents' goal and preferences in order to incorporate practical reasons in our analysis and to study their connection with epistemic reasons. This may also bring us closer to the cognitive trust model [19] where sentences like "i trusts j to do α in order to achieve φ" are dealt with.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: partial meet contraction and revision functions. J. Symbo. Logic **50**, 510–530 (1985)
2. Artemov, S.N.: The logic of justification. Rev. Symb. Logic **1**(4), 477–513 (2001)
3. Ayer, A.J.: Probability and Evidence. Columbia University Press, New York City (1972)
4. Baltag, A., Liu, F., Smets, S.: Reason-based belief revision in social networks. In: Slides, KNAW-Workshp on the Logical Dynamics of Information, Agency and Interaction, Amsterdam (2014)
5. Baltag, A., Moss, L.S., Solecki, S.: The logic of common knowledge, public announcements, and private suspicions. In: Gilboa, I. (ed.) Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 1998), pp. 43–56 (1998)
6. Baltag, A., Renne, B., Smets, S.: The logic of justified belief, explicit knowledge, and conclusive evidence. Ann. Pure Appl. Logic **165**(1), 49–81 (2014)

7. Baltag, A., Smets, S.: A qualitative theory of dynamic interactive belief revision. In: Wooldridge, M., Bonanno, G., van der Hoek, W. (eds.) Logic and the Foundations of Game and Decision Theory. Texts in Logic and Games, vol. 3. Amsterdam University Press, Amsterdam (2008)
8. van Benthem, J.: Dynamic logic for belief revision. J. Appl. Non Class. Logic **17**, 129–156 (2007)
9. van Benthem, J.: Logical Dynamics of Information and Interaction. Cambridge University Press, Cambridge (2011)
10. van Benthem, J., Fernández-Duque, D., Pacuit, E.: Evidence logic: a new look at neighborhood structures. In: Bolander, T., Braüner, T., Ghilardi, S., Moss, L. (eds.) Proceedings of Advances in Modal Logic, vol. 9, pp. 97–118. King's College Press, London (2012)
11. van Benthem, J., Pacuit, E.: Dynamic logics of evidence-based beliefs. Studia Logica **99**(1–3), 61–92 (2011)
12. Chellas, B.F.: Modal Logic: An Introduction. Cambridge University Press, Cambridge (1980)
13. Christoff, Z.: Dynamic logics of networks. Information Flow and the Spread of Opinion. Ph.D. thesis, ILLC, University of Amsterdam (2016)
14. Dastani, M., Herzig, A., Hulstijn, J., van der Torre, L.: Inferring trust. In: Leite, J., Torroni, P. (eds.) CLIMA 2004. LNCS, vol. 3487, pp. 144–160. Springer, Heidelberg (2005). doi:10.1007/11533092_9
15. Dietrich, F., List, C.: Judgment aggregation by quota rules: majority voting generalized. J. Theor. Polit. **19**(4), 391–424 (2007)
16. Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer, Berlin (2007)
17. Dubois, D., Liu, W., Ma, J., Prade, H.: The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. Inf. Fusion **32**, 12–39 (2016)
18. Dubois, D., Prade, H.: Representation and combination of uncertainty with belief functions and possibility measures. Comput. Intell. **4**(3), 244264 (1988)
19. Falcone, R., Castelfranchi, C.: Social trust: a cognitive approach. In: Tan, Y.-H., Castelfranchi, C. (eds.) Trust and Deception in Virtual Societies, pp. 55–90. Springer, Netherlands (2001). doi:10.1007/978-94-017-3614-5_3. Chap. 1
20. Friedman, J.: Suspended judgment. Philos. Stud. **162**(2), 165–181 (2013)
21. Grandi, U., Lorini, E., Perrussel, L.: Propositional opinion diffusion. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), pp. 989–997. ACM Press (2015)
22. Hunter, A., Booth, R.: Trust-sensitive belief revision. In: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015), pp. 3062–3068. AAAI Press (2015)
23. Jøsang, A.: A logic for uncertain probabilities. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **9**(3), 279–311 (2016)
24. Jøsang, A.: Interpretation of fusion and hyper opinions in subjective logic. In: 15th International Conference on Information Fusion, Singapore, pp. 1225–1232 (2017)
25. Keynes, J.M.: A Treatise on Probability. The Collected Writings, vol. 8. Macmillan, Hampshire (1973)
26. Kraus, S., Lehmann, D.J.: Knowledge, belief and time. Theoret. Comput. Sci. **58**, 155–174 (1988)
27. Liau, C.-J.: Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. Artif. Intell. **149**(1), 31–60 (2003)

28. Liu, F., Lorini, E.: Reasoning about belief, evidence and trust in a multi-agent setting (extended version). Technical report, Institut de Recherche en Informatique de Toulouse (IRIT) (2017)
29. Liu, F., Seligman, J., Girard, P.: Logical dynamics of belief change in the community. Synthese **191**(11), 2403–2431 (2014)
30. Lorini, E.: A minimal logic for interactive epistemology. Synthese **193**(3), 725–755 (2016)
31. Lorini, E., Jiang, G., Perrussel, L.: Trust-based belief change. In: Proceedings of the 21st European Conference on Artificial Intelligence (ECAI 2014), pp. 549–554. IOS Press (2014)
32. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
33. Singh, M.: Trust as dependence: a logical approach. In: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011), pp. 863–870. ACM (2011)
34. Xue, Y., Parikh, R.: Strategic belief updates through influence in a community. Stud. Logic **8**, 124–143 (2015)