Expressing Belief Flow in Assertion Networks

Sujata Ghosh[1,2] and Fernando R. Velázquez-Quesada[3]

[1] Department of Mathematics, Visva-Bharati, Santiniketan, West Bengal, India

[2] Center for Soft Computing Research, Indian Statistical Institute,
Kolkata, West Bengal, India

[3] Institute for Logic, Language and Computation, Universiteit van Amsterdam
Plantage Muidergracht 24, 1018 TV, Amsterdam, The Netherlands

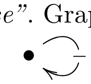
{S.Ghosh,F.R.VelazquezQuesada}@uva.nl

Abstract. In the line of some earlier work done on belief dynamics, we propose an abstract model of belief propagation on a graph based on the methodology of the revision theory of truth. A set of postulates is proposed, a dynamic language is developed for portraying the behavior of this model, and its expressiveness is discussed. We compare the proposal of this model with some of the existing frameworks for modelling communication situations.

Key words: assertion network, belief flow, belief merging, stability

1 Introduction

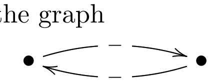
Self-reference is a tricky and complicated issue in logic. Ordinary propositional logic formulas can be expressed by trees, whereas there we have to resort to cyclic graphs (cf. [1]). In both cases, truth propagates backwards along the edges of finite trees or graphs. While this flow of truth stops in case of finite trees, giving a resultant truth value, it goes into a loop in case of cyclic graphs. Consider the *liar* statement *"this sentence is false"*. Graphically, it can be represented as



Gaifman's *pointer semantics* [2,3] and the *revision theory of truth* developed by Herzberger, Gupta and Belnap [4,5] provide semantics for sets of sentences with self-reference by looking at stable patterns among their truth values. Under these semantics, the value of the *liar* sentence never becomes stable as it oscillates between 1 and 0. On the other hand, for the *nested liars* sentences,

"The next sentence is false. The previous sentence is false."

which can be represented by the graph



there are two assignments, 1,0 and 0,1, that generate stable patterns under subsequent revisions of truth values. The main features of theories are the *backward propagation* of truth values along the edges (which correspond to the "revisions"), and the recognition of stable patterns.

In [6], the authors provide a formal model of real life communication situations using graphs where both *forward* and *backward* propagation of values are considered, which represents belief flow of a *reasoning agent*. This reasoning agent, also called the *observer*, wants to decide whether to believe or disbelieve certain facts, based on her and other agents' opinion about events/facts as well as agents. The observer's initial beliefs about the agents and facts are *revised* through an *iteration function* against the *merging* of rest of the information. A belief semantics via stability is defined, keeping the spirit of revision semantics mentioned earlier. [6] also provides a concrete model for such communication situations based on an infinite set of belief-values (a closed interval of real numbers). This gives rise to difficulties when segregating those values in terms of their interpretation and then studying their inter-dependence.

In order to overcome such difficulties and facilitate our formal modelling, we consider a finite set of belief-values here. Such finite sets play a significant role in better understanding of the underlying subtleties of the mutually conflicting opinions of the agents involved. We propose a set of postulates that a concrete model of the situations should satisfy.

To provide a sound formal foundation to our proposed model, we introduce a logical language to describe the revision process carried out. Instead of describing the outcome of the whole process (the general tradition of the logical approaches), we focus on the small-step dynamics of such situations, resembling the connectionist viewpoint. One of the main drawbacks of these approaches is the difficulty to provide an explanation of the underlying reasoning mechanism, viz. a *logical* description of the process, though attempts have been made to overcome it ([7,8,9,10]).

The main significance of this work lies in the fact that, though our model follow the connectionist framework, we have been able to provide a logical framework also so as to give a strong formal foundation to the proposed model. The search for an iteration function (the revision function which form an integral part of the model) that conforms to our intuitions is largely an empirical question. Still, our postulates impose basic restrictions on what this function should satisfy. They describe the way the observer's beliefs at a given stage will influence her beliefs after one step in the merging process.

The paper is organized as follows. In § 2, we recall the formal definition of the *Assertion Network Semantics* from [6] and propose several postulates stating properties the iteration function should satisfy. Then, we provide a concrete definition of such functions, and compare them with the postulates. We support our work with the aid of a software tool called *Assertion Network Toolkit*, which has been introduced in [6]. A dynamic logic of belief flow through this communication networks is proposed in § 3. Finally, § 4 focusses on comparison with some related works, with § 5 providing pointers towards future work.

2 Belief networks: a concrete model

In real life communication situations, we deal not only with sources of information with opinions about the facts/events, but also with opinions about each other. We can get information about the weather from a radio broadcasting, a webpage or a friend, and it is not strange to hear our friend saying *"you should not trust in those guys from the radio"*. Putting all this information together is not an easy task, but as highlighted in [6], the revision theoretic framework of Herzberger, Gupta, and Belnap [5,4] suggests a methodology that can be well applied in dealing with these rather complicated situations.

These situations can be represented by directed labelled graphs (DLG) with vertices representing facts and agents and edges representing agents' opinions. An edge labelled with "+" ("−") from a vertex n_1 to a vertex n_2 indicates that the agent represented by n_1 has positive (negative) opinion about the agent/fact represented by n_2. Although, in order to keep the model as simple as possible, we assign nodes to represent both *agents* and *facts*, we do differentiate them: agents are represented only with non-terminal nodes and facts with terminal ones. Agents with no opinions do not appear in the model.

An *external observer* reasons about the communication situations represented by the DLG. While the agents' opinions are represented by edges in the graph, the observer's beliefs are represented in the following way. Vertices are given values from a non-empty finite set Λ to indicate the states of belief of the observer regarding those agents and facts. As mentioned before, this is a departure from the models in [6], where the value set is a continuous interval, not the discrete set assumed here. We will see how this approach eventually aids in the understanding of the situation in a much more illuminating way and also provides a better insight into the language and logic of these networks.

Thinking of vertices of the graph as agents and facts rather than just sentences, leads from an analysis of truth as done in [4,5] to an analysis of a belief network. Consider the following example, given in [6].

> Suppose the observer is sitting in an office without windows. Next to her is her colleague (C), inside the same office. The observer is simultaneously talking on the phone to her friend (F), who is sitting in a street café.
>
> F: "*Everything your colleague says is false; the sun is shining!*"
> C: "*Everything your friend says is false; it is raining!*"

The information the observer has gathered can be described by the following graph where S is interpreted as *"the sun is shining"* and, while edges $\mathtt{F} \xrightarrow{+} \mathtt{S}$ and $\mathtt{C} \xrightarrow{-} \mathtt{S}$ represent the opinions the friend and the colleague have about S, edges $\mathtt{F} \xrightarrow{-} \mathtt{C}$ and $\mathtt{C} \xrightarrow{-} \mathtt{F}$ represent the opinions they have about each other.

$$\begin{tikzcd} \mathtt{F} \arrow[rr, bend left, "-"] \arrow[dr, "+"'] & & \mathtt{C} \arrow[ll, bend left, "-"] \arrow[dl, "-"] \\ & \mathtt{S} & \end{tikzcd}$$

Although there are two consistent truth value assignments, one of them is intuitively preferred, as the observer's friend has first hand experience of the weather in the street café. Based on this preference, the observer's beliefs *flow* through the graph: the contextually based stronger belief in `F` leads her to believe in `S`, but at the same time to disbelieve in `C`, since it is in conflict with `F`. Her disbelief in `C` in turn makes her belief in `S` stronger, which influences her belief in `F` once again. Both *forward* and *backward* propagation of beliefs are encountered.

This example shows both *backward* and *forward* propagation of beliefs. If a trusted source has some positive opinion about a certain proposition φ, the belief of the observer over φ will influence her belief on the trusted source, as well as the belief on the trusted source would have some effect over the observer's belief in φ. In the following, we try to base all these ideas on a more concrete level.

2.1 Assertion network semantics

An *Assertion Network Model* M is a tuple $M = (\mathcal{G}, \Psi)$, where

- $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \ell)$ is a directed labelled graph, with $\mathcal{V}$ the set of vertices, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ the set of edges and $\ell : \mathcal{E} \to \{+, -\}$ the labelling function.
- $\Psi : \Lambda^{\mathcal{V}} \to \Lambda^{\mathcal{V}}$ is the *iteration function*, with Λ the **set of values**.

Vertices represent agents and facts; edges represent agents' opinions.

The observer's beliefs are represented in a different way. We assume a function $H : \mathcal{V} \to \Lambda$, called an *hypothesis*, assigning to every vertex of $\mathcal{G}$ a value in Λ. The value $H(v)$ is interpreted as the state of belief the observer has about v.

The iteration function Ψ comes into play to combine forward and backward propagation, defining a **revision sequence** of the observer's beliefs. Given an initial hypothesis H, we define the sequence of functions $\langle H_i \,;\, i \in \omega \rangle$ as

$$H_0 := H, \qquad\qquad H_{i+1} := \Psi(H_i)$$

Inspired by the stability concept of revision theory, we can now define a partial stability semantics for our labelled graph. Let H be an initial hypothesis, v be a vertex in $\mathcal{V}$ and λ be a value in Λ. We say that λ **is the stable value of** v **starting from** H if there is $n \in \omega$ such that $H_i(v) = \lambda$ for all $i \geq n$. The **assertion network semantics** A_H is defined in this way:

$$\mathrm{A}_H(v) := \begin{cases} \lambda & \text{if } \lambda \text{ is the stable value of } v \text{ starting from } H \\ \text{undefined} & \text{if } \langle H_i(v) \,;\, i \in \omega \rangle \text{ oscillates.} \end{cases}$$

Following Theorem 1 and Theorem 2 in the section 2 of [6], it is pretty straightforward that,

Theorem 1. *The stable truth predicate of revision semantics is a special case of assertion network semantics, i.e., for every set of clauses Σ there is a labelled graph G and there are evaluation functions such that A_H coincides with the (partial) stable truth predicate on Σ.*[1]

[1] Here, we refer to a propositional language with *clauses* as described in [11], with the partial stable truth predicate defined in the proof of Theorem 2 in [6].

2.2 Postulates for the iteration function

As mentioned in the introduction, we will consider a finite set of belief-values for building up the *Assertion Network* model. We define the set as $\Lambda := \{-1, 0, 1\}$, where '-1' stands for disbelief, '0' for no opinion and '1' for belief.

The iteration function is the key of this model; it defines how the beliefs of the observer will be in the next stage, given her beliefs in the current one. Let us first make a brief analysis of what should be taken into account when deciding the next state of beliefs.

The case of facts is the simple one. To get her beliefs about some fact (represented by $v \in \mathcal{V}$) at stage $k+1$ ($H_{k+1}(v)$), the observer should take into account her current beliefs about the fact ($H_k(v)$) and her current beliefs about agents having an opinion (positive or negative) about the fact ($H_k(u)$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$). This is nothing but *forward propagation* of beliefs.

The case of an agent i (represented by $v \in \mathcal{V}$) is a more involved one. Besides her current beliefs about the agent ($H_k(v)$) and her current beliefs about agents having an opinion about i ($H_k(u)$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$; again, *forward propagation*), the observer should take into account the beliefs she has regarding agents and facts *about which i has an opinion* ($H_k(u)$ for every $u \in \mathcal{V}$ such that $\langle v, u\rangle \in \mathcal{E}$: *backward propagation*). All these will influence her next state of belief regarding the agent under consideration.

In the following, we propose some postulates for *rational iteration functions*. They reflect intuitive restrictions on how the belief state of the observer about some agent/fact should be modified during her introspection process.

Let $v \in \mathcal{V}$ be a terminal node (a fact) of the Assertion Network.

1. If **(a)** $H_k(u) = 1$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$+''$, and **(b)** $H_k(u) = -1$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$-''$, then $\Psi(H_k(v)) = H_{k+1}(v) = 1$ (the *positive enforcement of facts* postulate).
2. If **(a)** $H_k(u) = 1$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$-''$, and **(b)** $H_k(u) = -1$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$+''$, then $\Psi(H_k(v)) = H_{k+1}(v) = -1$ (the *negative enforcement of facts* postulate).
3. If **(a)** $H_k(u) = 0$ for every $u \in \mathcal{V}$ s.t. $\langle u, v\rangle \in \mathcal{E}$), then $\Psi(H_k(v)) = H_{k+1}(v) = H_k(v)$ (the *persistence of facts* postulate).

Now let $v \in \mathcal{V}$ be a non-terminal node (an agent).

1. If we have **(a)** and **(b)** from 1 of the terminal node case, plus **(c)** $H_k(u) = 1$ for every $u \in \mathcal{V}$ s.t. $\langle v, u\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$+''$, and **(d)** $H_k(u) = -1$ for every $u \in \mathcal{V}$ s.t. $\langle v, u\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$-''$, then $\Psi(H_k(v)) = H_{k+1}(v) = 1$ (the *positive enforcement of agents* postulate).
2. If we have **(a)** and **(b)** from 2 of the terminal node case, plus **(c)** $H_k(u) = 1$ for every $u \in \mathcal{V}$ s.t. $\langle v, u\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$-''$, and **(d)** $H_k(u) = -1$ for every $u \in \mathcal{V}$ s.t. $\langle v, u\rangle \in \mathcal{E}$ with $\ell\langle u, v\rangle =$ "$+''$), then $\Psi(H_k(v)) = H_{k+1}(v) = -1$ (the *negative enforcement of agents* postulate).
3. If we have **(a)** from 3 of the terminal node case, plus **(b)** $H_k(u) = 0$ for every $u \in \mathcal{V}$ s.t. $\langle v, u\rangle \in \mathcal{E}$, then $\Psi(H_k(v)) = H_{k+1}(v) = H_k(v)$ (the *persistence of agents* postulate).

2.3 Concrete model

We provide a concrete definition of the iteration function, describing the change in observer's beliefs about an agent/fact depending on those of related ones.

Let $M = (\mathcal{G}, \Psi)$ be an *Assertion Network Model* with $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \ell)$. For a vertex $v \in \mathcal{V}$, define

$$\begin{aligned} \mathrm{In}^+(v) &:= \{w \in \mathcal{V}\,;\, \ell\langle w, v\rangle = \text{“}+\text{”}\}, & \mathrm{In}^-(v) &:= \{w \in \mathcal{V}\,;\, \ell\langle w, v\rangle = \text{“}-\text{”}\}, \\ \mathrm{Out}^+(v) &:= \{w \in \mathcal{V}\,;\, \ell\langle v, w\rangle = \text{“}+\text{”}\}, & \mathrm{Out}^-(v) &:= \{w \in \mathcal{V}\,;\, \ell\langle v, w\rangle = \text{“}-\text{”}\} \end{aligned}$$

The set $\mathrm{In}(v) := \mathrm{In}^+(v) \cup \mathrm{In}^-(v)$ consists of the vertices that *can reach* v. The set $\mathrm{Out}(v) := \mathrm{Out}^+(v) \cup \mathrm{Out}^-(v)$ consists of the vertices that *can be reached from* v. The set of terminal vertices of $\mathcal{G}$ can be defined as $\mathcal{T}_{\mathcal{G}} := \{v \in \mathcal{V}\,;\, \mathrm{Out}(v) = \emptyset\}$.

Let H be an hypothesis. For every $w \in \mathrm{In}(v)$, define s^v_w as the H-value of w with sign according to the label of the edge that links it with v; for every $w \in \mathrm{Out}(v)$, define t^v_w as the H-value of w with sign according to the label of the edge that links v to it. Formally,

$$s^v_w := \begin{cases} H(w) & \text{if } w \in \mathrm{In}^+(v) \\ -H(w) & \text{if } w \in \mathrm{In}^-(v) \end{cases} \qquad t^v_w := \begin{cases} H(w) & \text{if } w \in \mathrm{Out}^+(v) \\ -H(w) & \text{if } w \in \mathrm{Out}^-(v) \end{cases}$$

For each value $\lambda \in \Lambda$, define S^v_λ as the set of vertices $w \in \mathrm{In}(v)$ such that $s^v_w = \lambda$; similarly, define T^v_λ as the set of vertices in $w \in \mathrm{Out}(v)$ such that $t^v_w = \lambda$.

$$S^v_\lambda := \{w \in \mathrm{In}(v)\,;\, s^v_w = \lambda\} \qquad T^v_\lambda := \{w \in \mathrm{Out}(v)\,;\, t^v_w = \lambda\}$$

For a terminal vertex $v \in \mathcal{T}_{\mathcal{G}}$, its $\Psi(H)$-value depends on the H-values of v itself and on those of the vertices in $\mathrm{In}(v)$. Here is our particular definition.

$$\Psi(H)(v) := \begin{cases} 1 & \text{if } |S^v_1| > |S^v_{-1}| \\ -1 & \text{if } |S^v_1| < |S^v_{-1}| \\ H(v) & \text{otherwise.} \end{cases}$$

For a non-terminal vertex $v \in \mathcal{V} \backslash \mathcal{T}_{\mathcal{G}}$, the definition is a bit more complicated. Unlike the terminal ones, in addition to the current value of v we now have to take into account the influences of both the incoming edges as well as the outgoing ones, since we want to represent both *forward* and *backward* propagation of beliefs. The value suggested by the incoming edges (IE_v) and the one suggested by the outgoing ones (OE_v) are considered separately.

$$IE_v := \begin{cases} 1 & \text{if } |S^v_1| > |S^v_{-1}| \\ -1 & \text{if } |S^v_1| < |S^v_{-1}| \\ H(v) & \text{otherwise.} \end{cases} \qquad OE_v := \begin{cases} 1 & \text{if } |T^v_1| > |T^v_{-1}| \\ -1 & \text{if } |T^v_1| < |T^v_{-1}| \\ H(v) & \text{otherwise.} \end{cases}$$

Their combination gives the $\Psi(H)$-value of v defined by the following table:

$IE_v \backslash OE_v$	-1	0	1
-1	-1	-1	0
0	-1	0	1
1	0	1	1

With this definition of Ψ, the next theorem can be easily proved.

Theorem 2. *Ψ satisfies the three fact postulates and the three agent postulates.*

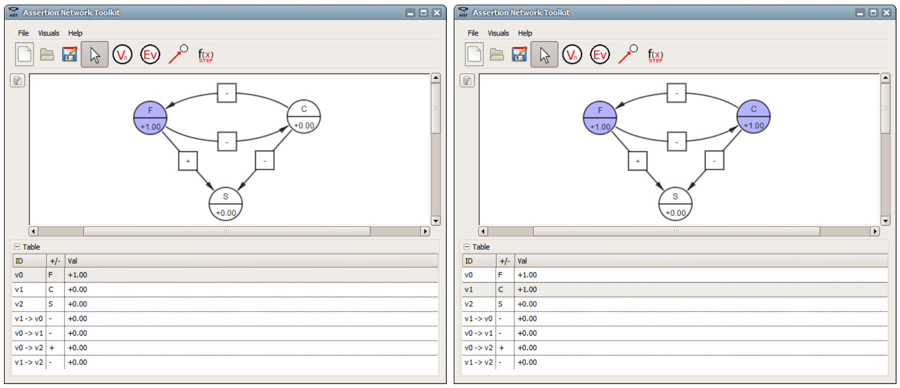


Fig. 1. Initial opinions

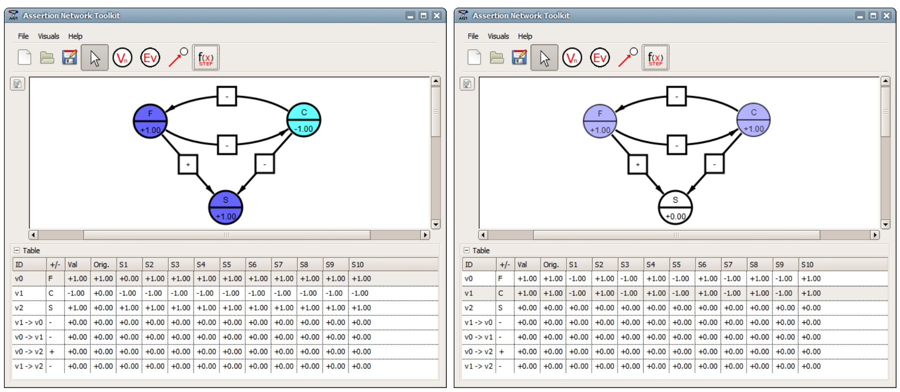


Fig. 2. Final opinions

2.4 Assertion network toolkit

As mentioned in the introduction, looking for the adequate iteration function Ψ is largely an empirical task. We can claim that the definition given just now is a plausible one since it satisfies all the postulates, but still more complicated examples are to be checked to validate the claim that this particular definition reflects our intuitive interpretation. The *Assertion Network Toolkit* (ANT), presented in [6], allows us to play around with the functions and values.

As an example of its use, consider the communication situation described at the beginning of this section. The iteration function Ψ defined in the earlier subsection is the one currently implemented in the ANT. Figure 1 shows two screenshots with difference in the values of the initial hypothesis. Figure 2 shows the corresponding final values after the iteration process.

In the first case (the left hand side of Figure 1), the observer believes in her friend because of the friend's first hand experience of what is happening outside the office; besides that, she does not have any initial opinion about her colleague or the discussed fact. In this setting, the initial hypothesis H_0 assigns a value of 1 to the vertex representing the friend ($H_0(\mathtt{F}) = 1$) and a value of 0 to the others ($H_0(\mathtt{C}) = H_0(\mathtt{S}) = 0$). In the second case (the right hand side of Figure 1), the observer has an equally high initial opinion about her friend as well as her colleague. The initial hypothesis H_0 assigns a value of 1 to both her friend and colleague ($H_0(\mathtt{C}) = H_0(\mathtt{F}) = 1$), but 0 to the mentioned fact ($H_0(\mathtt{S}) = 0$). We let the program iterate the function several times, getting the screenshoots of Figure 2 and the sequence of values of the tables below.

	H_0	H_1	H_2	H_3	$\cdots$
F	1	1	1	1	$\cdots$
C	0	-1	-1	-1	$\cdots$
S	0	1	1	1	$\cdots$

	H_0	H_1	H_2	H_3	$\cdots$
F	1	-1	1	-1	$\cdots$
C	1	-1	1	-1	$\cdots$
S	0	0	0	0	$\cdots$

In the first case, all the vertices reach stable values (in just two steps); in the second case, only the vertex representing `S` gets a stable value: that of *"no opinion"* of the observer (the values of `F` and `C` oscillate). The readers will definitely consent to the fact that in both these cases, the final belief values completely agree with our intuitions.

3 Expressing belief networks

This section provides a logical language to express the behavior of the Assertion Network Model. The network focuses on the observer's point of view, so we define a language that takes her perspective. The atomic propositions are expressions indicating the state of belief the observer has about agents or facts portrayed in the network, and then we build more complex formulas using the standard logical connectives. This language does not describe the graph (we cannot express things like *"agent i has a positive opinion about p"*), but it describes the observer's beliefs about the represented situation; this will serve our purpose here. For readers interested in a more expressive language, we refer to [12].

In the language we provide a way to talk about the most important part of the model: the update of beliefs carried out by the iteration function. We introduce the syntactic operator $\bigcirc$ to represent the iteration function: it allows us to talk about what happens with the observer's beliefs after one step of revision and merging of beliefs. Formulas of the form $\bigcirc\varphi$ are read as *"after one iteration of the function, φ is the case"*. This operator describes the way the beliefs of the reasoning agent propagate through the network after a single iteration step.

Finally, we are also interested in the outcome of the whole process. Such a process reaches an end whenever the beliefs of the observer become *stable*, that is, whenever they reach a stage from which further iterations of the function will not change them anymore (which is not always the case). We introduce the syntactic operator $\circledast$; it represents the stable stage reached by the network (whenever it exists) and allows us to talk about what happens with the observer's belief at the end of the process (if it ever ends). Formulas of the form $\circledast\varphi$ are read as *"after the whole process, φ is the case"*.

3.1 A language expressing the observer's beliefs

Given a set of *agents* $\mathbf{A}$ and a set of *propositions* $\mathbf{P}$, the *Language of Beliefs* $\mathcal{LB}$ is given by:

$$\varphi := \mathsf{B}\gamma \mid \mathsf{N}\gamma \mid \mathsf{D}\gamma \mid \neg\varphi \mid \varphi \vee \psi$$

with $\gamma \in \mathbf{A} \cup \mathbf{P}$. Formulas of the form $\mathsf{B}\gamma$ indicates *"the observer believes in γ"*, while $\mathsf{N}\gamma$ indicates *"the observer does not have any opinion about γ"* and $\mathsf{D}\gamma$ indicates *"the observer disbelieves in γ"*.

To avoid any confusion that may arise due to the use of the traditional intensional operators in an extensional language, we make the following remarks.

- Formulas in $\mathcal{LB}$ express exclusively *the observer's* beliefs.
- The language $\mathcal{LB}$ *is not a modal language*. Its atomic propositions $\mathsf{B}\gamma$, $\mathsf{N}\gamma$ and $\mathsf{D}\gamma$ have special meanings, but they are still atomic propositions.
- Usually, the truth values of atomic propositions are not related in any way. Here, the semantics will be defined in a way that the truth values of some of them are related: formulas like $\mathsf{B}\gamma \wedge \mathsf{D}\gamma$, for example, will never be true.

Formulas of $\mathcal{LB}$ are interpreted in Assertion Network Models by assuming a map that uniquely identify each vertex of the model with an agent or a fact in $\mathbf{A} \cup \mathbf{P}$. The map should satisfy our requirement: facts have to be mapped to terminal vertices.

Let $M = (\mathcal{G}, \Psi)$ be an Assertion Network Model, with $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \ell)$. An *interpretation* I is a partial injective function $I : \mathbf{A} \cup \mathbf{P} \to \mathcal{V}$ such that, for each $p \in \mathbf{P}$, we have $I(p) \in \mathcal{T}_{\mathcal{G}}$, when it is defined. Given I and an initial hypothesis H, the truth definition of formulas of $\mathcal{LB}$ in M is given by

$$\begin{array}{lll}
M, I, H \models \mathsf{B}\gamma & \text{iff} & H(I(\gamma)) = 1 \\
M, I, H \models \mathsf{N}\gamma & \text{iff} & H(I(\gamma)) = 0 \\
M, I, H \models \mathsf{D}\gamma & \text{iff} & H(I(\gamma)) = -1 \\
M, I, H \models \neg\varphi & \text{iff} & M, I, H \not\models \varphi \\
M, I, H \models \varphi \vee \psi & \text{iff} & M, I, H \models \varphi \;\text{ or }\; M, I, H \models \psi
\end{array}$$

Thus, the formula $\mathsf{B}\gamma$ (resp. $\mathsf{N}\gamma$, $\mathsf{D}\gamma$) is true in the model M under the interpretation I if and only if the H-value of the graph component to which γ is mapped is equal to 1 (resp. 0, -1).

3.2 A language expressing belief flow

The language $\mathcal{LB}$ is static, in the sense that it does not express how beliefs change as a result of the belief propagation. Here, we extend the language with two dynamic operators that allows us to talk about the model after one iteration step ($\bigcirc$) and also after it reaches a stable situation ($\circledast$). The full language of the *Logic of Belief Flow* $\mathcal{LBF}$, is given by:

$$\varphi := \mathsf{B}\gamma \mid \mathsf{N}\gamma \mid \mathsf{D}\gamma \mid \neg\varphi \mid \varphi \vee \psi \mid \bigcirc\varphi \mid \circledast\varphi$$

with $\gamma \in \mathbf{A} \cup \mathbf{P}$. Formulas of the form $\bigcirc\varphi$ express *"after the observer once considers the information she has, φ is the case"*; formulas of the form $\circledast\varphi$ express *"after the observer considers all the information she has, φ is the case"*.

While the operator $\bigcirc$ represents one step in the iteration process, the operator $\circledast$ represents stable positions. The first one gets truth value by using the iteration function Ψ; the second one looks for iterations that do not change values from some moment on.

$$\begin{array}{lll} M, I, H \models \bigcirc\varphi & \text{iff} & M, I, \Psi(H) \models \varphi \\ M, I, H \models \circledast\varphi & \text{iff} & \exists n \in \omega \text{ such that } M, I, \Psi^i(H) \models \varphi \text{ for all } i \geq n. \end{array}$$

To close this section, we give examples of formulas that hold in the Assertion Network Model corresponding to the example described in § 2 and whose iterated values are shown in tables of page 8. Formally, we have

– $\mathcal{V} := \{\mathtt{F}, \mathtt{C}, \mathtt{S}\}$; $\mathcal{E} := \{\langle \mathtt{F}, \mathtt{C}\rangle, \langle \mathtt{C}, \mathtt{F}\rangle, \langle \mathtt{F}, \mathtt{S}\rangle, \langle \mathtt{C}, \mathtt{S}\rangle\}$,
– $\ell\langle \mathtt{F}, \mathtt{S}\rangle =$ "+"; $\ell\langle \mathtt{F}, \mathtt{C}\rangle = \ell\langle \mathtt{C}, \mathtt{F}\rangle = \ell\langle \mathtt{C}, \mathtt{S}\rangle =$ "−".

and Ψ as defined before. The initial hypothesis H is given by

$$H(\mathtt{F}) = 1 \qquad H(\mathtt{C}) = 0 \qquad H(\mathtt{S}) = 0$$

From the values shown in the corresponding table, we have that the following formulas hold in M, I, H:

- $\mathsf{B}\mathtt{F} \wedge \mathsf{N}\mathtt{C} \wedge \mathsf{N}\mathtt{S}$
- $\bigcirc(\mathsf{B}\mathtt{F} \wedge \mathsf{D}\mathtt{C} \wedge \mathsf{B}\mathtt{S})$
- $\bigcirc\bigcirc(\mathsf{B}\mathtt{F} \wedge \mathsf{D}\mathtt{C} \wedge \mathsf{B}\mathtt{S})$
- $\circledast(\mathsf{B}\mathtt{F} \wedge \mathsf{D}\mathtt{C} \wedge \mathsf{B}\mathtt{S})$

Considering some variations of the initial hypothesis, ANT shows us that the following formulas also hold.

- $(\mathsf{B}\mathtt{F} \wedge \mathsf{B}\mathtt{S}) \rightarrow \circledast(\mathsf{B}\mathtt{F} \wedge \mathsf{D}\mathtt{C} \wedge \mathsf{B}\mathtt{S})$
  If the observer initially believes in F *and* S*, then her initial belief about* C *is irrelevant.*
- $(\mathsf{B}\mathtt{F} \wedge \mathsf{B}\mathtt{C} \wedge \mathsf{N}\mathtt{S}) \rightarrow ((\bigcirc^k\mathsf{B}\mathtt{F} \rightarrow \bigcirc^{k+1}\mathsf{D}\mathtt{F}) \wedge (\bigcirc^k\mathsf{D}\mathtt{F} \rightarrow \bigcirc^{k+1}\mathsf{B}\mathtt{F})) \quad (k \geq 0)$
  If she initially believes in F *and* C *without having an opinion about* S*, then her beliefs about* F *will oscillate (*$\bigcirc^0\varphi := \varphi$ *and* $\bigcirc^{k+1}\varphi := \bigcirc\bigcirc^k\varphi$*).*
- $(\mathsf{B}\mathtt{F} \wedge \mathsf{B}\mathtt{C} \wedge \mathsf{N}\mathtt{S}) \rightarrow \neg\circledast(\mathsf{B}\mathtt{F} \vee \mathsf{N}\mathtt{F} \vee \mathsf{D}\mathtt{F})$
  Therefore, there is no stable value for F*.*
- $(\mathsf{B}\mathtt{F} \wedge \mathsf{B}\mathtt{C} \wedge \mathsf{N}\mathtt{S}) \rightarrow \circledast\mathsf{N}\mathtt{S}$
  But there is a stable value (viz. 0) for S*.*

Evidently, the last three formulas express the observer's opinions in the second example we dealt with in Section 2.4. Finally, we also have some validities which provide some insights towards the complete axiomatization of the proposed logic, which we leave for future work:

- $\mathsf{B}\gamma \rightarrow (\neg\mathsf{N}\gamma \wedge \neg\mathsf{D}\gamma)$
- $\mathsf{D}\gamma \rightarrow (\neg\mathsf{N}\gamma \wedge \neg\mathsf{B}\gamma)$
- $\mathsf{N}\gamma \rightarrow (\neg\mathsf{B}\gamma \wedge \neg\mathsf{D}\gamma)$
- $\bigcirc(\varphi \wedge \psi) \leftrightarrow (\bigcirc\varphi \wedge \bigcirc\psi)$
- $\circledast(\varphi \wedge \psi) \leftrightarrow (\circledast\varphi \wedge \circledast\psi)$
- $\circledast\varphi \rightarrow \circledast\bigcirc\varphi$

4 Other models and logics: a comparison

An extensive amount of work has been done in formalizing and modelling the revising/merging of belief/information. Here, we provide a discussion to compare our approach with a few of the existing ones.

4.1 Different approaches for revising/merging

The classical work on belief revision, the *AGM approach* ([13]), introduces postulates that an operator performing revision should satisfy in order to be considered rational. Several other frameworks have been proposed; particularly related with our proposal are those focused on *iterated* revision, like [14] and [15]. The field has extended to incorporate the more general branch of *belief merging*, focussed on situations where both the current and the incoming information have the same priority and the same structure ([16,17,18]).

Our approach lies on the revision side, with the agents' opinions and the observer's beliefs being represented in a different way. Nevertheless, we do not consider simple revision, but *revision by merging*, since the observer revises her beliefs against the merged opinions of all the agents involved, very much in the spirit of [19]. Also, the main novelty of our work is that it considers agents that have opinions not only about the discussed facts, but also about themselves.

The dynamic logic framework provides a natural way to express changes in information. Various logics have been proposed, like dynamic epistemic logic (*DEL*; [20,21]) and dynamic doxastic logic (*DDL*; [22]). In [23] the author looks into *DEL* and *AGM* belief revision, providing a joint perspective. While *DDL* captures the *AGM* postulates for belief revision in a logical language, *DEL* talks about concrete information update procedures that change models/situations.

In contrast, $\mathcal{LBF}$ focusses on introspection of a reasoning agent regarding the transition of her belief states in a communication situation. Belief states are expressed in a propositional language, and their transition is captured by the dynamic modal operators $\bigcirc$ and $\circledast$. Note how *DDL* expresses agents' beliefs after a certain revision process that occur in her doxastic state, while *DEL* provides a framework for dealing with *hard* information (changing the knowledge state) as well as *soft* information (affecting beliefs). $\mathcal{LBF}$ is proposed to capture the process of continuing change in the opinions/beliefs that goes on in the observer's mind in the described situations.

On the other side of the spectrum, and closer to the Assertion Network semantics, there are approaches based on interconnected networks, where the results of the process may sometime corroborate with the stability concepts, and in some other cases, have quite different approaches, e.g. the probabilistic one. To mention a few, in [24,25], a *Neural-Logic Belief Network* (*NLBN*, a neuro-symbolic network) is defined which can be used to model common sense reasoning, providing a way for representing changes in the agent's belief attitudes. In [26], the authors propose a distributed approach to belief revision, where numerical values as probability measures have been incorporated in the models to represent degrees of uncertainty, and computations are performed using Dempster rule and Bayesian conditioning.

We should also mention *Bayesian Belief Nets* (*BBN*; [27]) in this regard. They are directed acyclic graphs, with nodes representing variables and the edges representing their causal relationship. The value of a variable is given by a conditional probability table, based on the causal relationship calculated with Bayes' rule. Based on these tables, *BBN* can be used in decision making, where both inductive as well as deductive reasoning can be performed.

Let us compare those approaches with the model of § 2. The novelty lies in the semantics derived from stability as used in the revision theory of truth [5,4]. In *NLBN*, only forward propagation is considered and the representation is restricted to propositions, while our model considers backward propagation and represents agents as well. Similar is the case of *BBN*, though in some sense their probabilistic approach can be used in a greater variety of domains. The work of [26] is closer to ours, though with subtle but important differences, the most notable among them being our very centralized approach.

Different from connectionist approaches, logical ones have the advantage of providing a better understanding of the underlying process. On the other hand, networks and the stability concept are natural representations of the interconnected information and the discussion process that leads to agreements. In [19], the authors propose a combination: conciliation via iterated belief merging. Beliefs of several agents are merged, and then each one of them revises her own beliefs with respect to the result of the merging. The process is repeated until a fixed point is reached; the conciliation operator is defined with respect to it.

As in our work, they look for *stable* situations, where further interaction between the diverse components will not modify the current status. Somewhat similar to our approach, they use a two-stage iterative process: merging and then revising. But, once again, in this work as well as in similar such, the basic focus lies on different agents' belief sets with no mention of belief/trust over other agents, where the novelty of our work lies.

4.2 Small steps

The idea of focusing on the small steps of a process is not new. It has been a proposed solution for the so called *logical omniscience* problem, about unrealistic assumptions on the agents' reasoning power.

In [28,29], Duc proposes a dynamic epistemic logic to reason about agents that are neither logically omniscient nor logically ignorant. The main idea is to represent the knowledge of an agent as a set of formulas, and to allow her to improve her information set as time goes by. Instead of representing agents that know everything from the very beginning, this approach focusses on the step-by-step process that leads to that outcome. Our work shares this concept: we focus on the small steps in belief revision/merging process. In some cases, the small steps will lead to stable values, indicating that the (possibly inconsistent) initial information and the observer's initial beliefs can be merged. In others, the values will oscillate, indicating that they cannot find a way to live together.

4.3 Trust

One of the main features of the Assertion Networks is that it allows us to represent not only opinions about facts, but also opinions about agents. This can be interpreted as the observer's *trust*, allowing us to represent asymmetries in the way the agents' opinions will influence the observer's beliefs. Several works have analyzed the notion of *trust* in multi-agent systems.

In [30], Liau proposes a modal logic with three operators: B_i (*"i believes φ"*), I_{ij} (*"j informs i about φ"*) and T_{ij} (*"i trust j about φ"*). Beliefs and information are normal modal operators, so an agent's beliefs are closed under logical consequence, and once she acquires some information from another agent, she also acquires all its logical consequences. Trust, on the other hand, is given by an operator with neighborhoods semantics, so trusting another agent about φ does not make an agent to trust her about the logical consequences of φ.

In [31], the authors extend Liau's work by introducing *topics* and *questions*. As they observe, Liau's work explains the consequence of trust, but does not explain where trust comes from. The notion of *topic* allows to create trust of agent i on agent j about fact ψ whenever i trusts j about a fact φ that shares the same topic with ψ (*topic-based trust transfer*). The notion of *question* allows to create trust or distrust from the answer of some question (*question-based trust derivation* and *question-based distrust derivation*).

In our proposal, the notion of belief in an agent, different from the notion of trust of the described works, is not relative to a particular statement (as formulas of the form $T_{ij}\varphi$ express), but relative to the agent itself. Also, since facts are represented independently from each other, beliefs of the observer are not closed under any inference relation. Moreover, the observer's initial beliefs about the facts and the agents are not necessarily related: the agent can initially believe in p without believing in agents having a positive opinion about p.

The described approaches work on a static level, without considering *dynamics* of the system. Even exchanges of information and questions are semantically represented as properties of the model, and not as actions that modifies it. The main focus of our approach is the *dynamic* process through which all the involved participants interact, updating the model and influencing themselves while trying to reach an agreement.

5 Conclusion and intentions

In this work, we propose a model of belief propagation based on the methodology of the revision theory of truth. A dynamic language is developed for expressing the belief flow in the model in terms of an external observer's introspection process. We have compared the model and the language with some of the existing frameworks for modelling communication situations.

In our framework, the next-stage belief value of a node is given in terms of the current beliefs about the incoming and outgoing nodes (forward and backward propagation). Our postulates state the behaviour of the iteration function in completely biased cases. Some further avenues of investigation are as follows.

Different iteration functions. In more general situations, there is no unique way to define the iteration function. A *majority-based* one may represent an observer that follows the majority, a *confident* one can represent an agent that gives more weight to her current beliefs and a *credulous* one can represent observers that give more precedence to others' opinions. It will be interesting to formalize these different policies.

Opinionated edges. We can consider beliefs not only about facts and agents, but also about the opinions. We can think of situations where an agent is an expert in some subject but not in some other. Thus in some cases it is more natural to have different degrees of beliefs in the agent's different opinions.

Extending value set. We have considered a three-valued belief-degree set Λ here, but it can be easily extended to any finite valued one, so as to express more possible epistemic states of the observer. The model will get closer to the actual real life situations.

Comparing expressivity. In the presented language, formulas of the form $\circledast\varphi$ express stable values, related with fixed points in some sense. It would be interesting to make a study about the expressiveness of the language compared with fixpoint logics, like the modal μ-calculus.

Acknowledgements. We gratefully acknowledge Erik Scorelle for working on *ANT* with the modified definitions and helping in with the screenshots. We also thank Benedikt Löwe for the initial discussions that provided important pointers towards the definition of the model, and for going through a preliminary draft of this paper, providing insightful comments. Finally, we thank the editors and anonymous referees for their useful suggestions and comments. The second author acknowledges a scholarship by **CONACyT** (# 167693) from México.

References

1. Bolander, T.: Self-reference and logic. Phi News **1** (2002) 9–44
2. Gaifman, H.: Operational pointer semantics: Solution to self-referential puzzles I. In Vardi, M., ed.: Proceedings of TARK 1988. (1988) 43–59
3. Gaifman, H.: Pointers to truth. Journal of Philosophy **89** (1992) 223–261
4. Gupta, A., Belnap, N.D.: The Revision Theory of Truth. The MIT Press, Cambridge, MA (1993)
5. Herzberger, H.G.: Naive semantics and the liar paradox. The Journal of Philosophy **79** (1982) 479–497
6. Ghosh, S., Löwe, B., Scorelle, E.: Belief flow in assertion networks. In Priss, U., Polovina, S., Hill, R., eds.: Proceedings of ICCS 2007, Sheffield, UK. Volume 4604 of LNAI., Heilderberg, Springer-Verlag (July 2007) 401–414
7. Balkenius, C., Gärdenfors, P.: Nonmonotonic inferences in neural networks. In Allen, J.A., Fikes, R., Sandewall, E., eds.: Proceedings of KR'91, (1991) 32–39
8. Blutner, R., Doherty, P.D.: Nonmonotonic logic and neural networks. Available at http://citeseer.ist.psu.edu/160749.html
9. d'Avila Garcez, A., Broda, K., Gabbay, D.: Symbolic knowledge extraction from trained neural networks: A sound approach. Artificial Intelligence **125**(1–2) (2001) 155–207

10. Leitgeb, H.: Interpreted dynamical systems and qualitative laws: from neural networks to evolutionary systems. Synthese **146**(1-2) (August 2005) 189–202
11. Löwe, B.: Revision forever! In Schärfe, H., Hitzler, P., Øhrstrøm, P., eds.: Proceedings of ICCS 2006, Aalborg, Denmark. Volume 4068 of LNAI., Berlin, Springer-Verlag (July 2006) 22–36
12. Ghosh, S., Velázquez-Quesada, F.R.: Merging information, working paper.
13. Alchourrón, C., Gärdenfors, P., Makinson, D.: On the logic of theory change. The Journal of Symbolic Logic **50** (1985) 510–530
14. Darwiche, A., Pearl, J.: On the logic of iterated belief revision. In Fagin, R., ed.: Proceedings of TARK 1994, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc. (1994) 5–23
15. Jin, Y., Thielscher, M.: Iterated belief revision, revised. Artificial Intelligence **171**(1) (2007) 1–18
16. Konieczny, S.: Sur la logique du changement: révision et fusion de bases de connaissance. PhD thesis, Laboratoire d'Informatique Fondamentale de Lille (1999)
17. Konieczny, S., Pino-Pérez, R.: On the logic of merging. In: Proceedigns of KR'98. (1998) 488–498
18. Konieczny, S., Pino-Pérez, R.: Merging information under constraints: a logical framework. Journal of Logic and Computation **12**(5) (2002) 773–808
19. Gauwin, O., Konieczny, S., Marquis, P.: Conciliation through iterated belief merging. Journal of Logic and Computation **17**(5) (October 2007) 909–937
20. Gerbrandy, J.: Bisimulations on Planet Kripke. PhD thesis, Institute for Logic, Language and Computation (University of Amsterdam) (1999)
21. Plaza, J.A.: Logics of public communications. In Emrich, M.L., Pfeifer, M.S., Hadzikadic, M., Ras, Z.W., eds.: Proceedings of ISMIS 1989. (1989) 201–216
22. Segerberg, K.: The basic dynamic doxastic logic of AGM. In Williams, M., Rott, H., eds.: Frontiers in Belief Revision, Kluwer, Dordrecht (2001) 57–84
23. van Benthem, J.: Dynamic logic for belief revision. Journal of Applied Non-Classical Logic **17**(2) (2007) 129–155
24. Low, B., Foo, N.: Towards a network for representing beliefs. In: Proceedings of the 2nd International Computer Science Conference. (1992) 85–91
25. Low, B., Foo, N.: A network formalism for commonsense reasoning. In: Proceedings of the Sixteenth Australian Computer Science Conference. (1993) 425–434
26. Dragoni, A., Giorgini, P.: Distributed belief revision. Autonomous Agents and Multi-Agent Systems **6**(2) (2003) 115–143
27. Williamson, J.: Bayesian nets and causality, Philosophical and computational foundations. OUP (2005)
28. Duc, H.N.: Logical omniscience vs. logical ignorance on a dilemma of epistemic logic. In: Proceedings of EPIA'95, London, UK, Springer-Verlag (1995) 237–248
29. Duc, H.N.: Reasoning about rational, but not logically omniscient, agents. Journal of Logic and Computation **7**(5) (1997) 633–648
30. Liau, C.J.: Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. Artificial Intelligence **149**(1) (2003) 31–60
31. Dastani, M., Herzig, A., Hulstijn, J., van der Torre, L.W.N.: Inferring trust. In Leite, J.A., Torroni, P., eds.: CLIMA V. Volume 3487 of LNCS., Springer (2004) 144–160