

Expressing Belief Flow in Assertion Networks

Sujata Ghosh^{1,2} and Fernando R. Velázquez-Quesada³ *

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 modal language is developed for portraying the behavior of this model, and its expressiveness is discussed. We compare the proposal of this model as well as the language developed with some of the existing frameworks for modelling communication situations.

1 Introduction

Self-reference is a very tricky and complicated issue in logic. Ordinary propositional logic formulas can be expressed by trees, whereas there we have to resort to cyclic graphs ([4] provides a detailed discussion on this issue). In both cases, truth propagates backwards along the edges of the 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. For example, consider the *liar* statement: *This sentence is false*. Graphically, it can be represented as,



Gaifman Pointer semantics [11,12] and the revision theory of truth developed by Herzberger, Gupta and Belnap [17,18] provide semantics for sets of sentences with self-reference by looking at stable patterns among the truth values of the

* Acknowledges a scholarship by **Consejo Nacional de Ciencia y Tecnología (CONACyT)** from México. Scholarship holder # 167693.

sentences in the set. Under these semantics, the value of the *liar* sentence never becomes stable since 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 this revision theory are the *backward propagation* of truth values along the edges (which correspond to the “revisions”), and the recognition of stable patterns.

In [15], the authors provide a formal model of real life communication situations using graphs where both *forward* and *backward* propagation of values along edges are considered. In this case, the propagation represents the way the beliefs of a *reasoning agent* flows. This reasoning agent (also called the *observer*) wants to decide whether to believe or disbelieve certain facts, and her only available information is opinions of other agents about those facts. The situation becomes more interesting when the individual agents also have opinions about each other. The observer has some initial beliefs about the agents and facts; those beliefs are *revised* against the other information.

In the proposed framework, the *Assertion Network Semantics*, the observer’s information is represented as a directed labelled graph with nodes corresponding to agents or facts and edges corresponding to opinions. The observer’s beliefs are represented by values assigned to the nodes; these values are iteratively updated through an *iteration function* that represents the step-by-step revising/merging process. A belief semantics via stability is defined while keeping the spirit of revision semantics mentioned earlier.

The set of belief-values considered in [15] while providing a concrete model for such communication situations is infinite, viz. a closed interval of real numbers. A continuous set of such values gives rise to certain difficulties, eg. segregating those values in terms of their interpretation, as well as studying their interdependence, which forms the basic focus of our approach, given the situations we intend to model.

With a goal towards overcoming such difficulties and facilitate the formal modelling of such communication situations, in the present work we consider a finite set of belief-values. Such finite sets play a significant role in the better understanding of the underlying subtleties of the mutually conflicting opinions of the agents involved. To provide a sound formal foundation to our proposed model, we propose a logical language to describe the revision process it carries out. Instead of just describing the outcome of the whole process, which has been the general tradition of the logical approaches, we channel our interests in the small-step dynamics of such situations, resembling the connectionist viewpoint.

In line with the proposals for describing the underlying process in connectionist approaches to reasoning, revising current states of beliefs as well as merging information and beliefs [2,3,7,23], we are interested in how the observer uses the information she currently has so as to reach a state of mind in which she finally has firm beliefs about the agents and the facts involved. It should be mentioned here that one of the main drawbacks of these connectionist approaches is the difficulty to provide an explanation of the underlying reasoning mechanism, that is, the difficulty to provide a *logical* description of the process. Some attempts have been made to overcome it in [7]. Also, [2,3,23] provide some interesting links between non-monotonic reasoning and the underlying propagation mechanism in their neural network models.

The main significance of this work lies in the fact that, though our model of communication situations 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. As has been pointed out by the authors in [15], 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, we can impose certain basic restrictions on what this function should satisfy. We propose some postulates that 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* and propose several postulates stating properties the iteration function should satisfy. Then, we provide a concrete definition of such function, 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 [15]. 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 some pointers towards future work.

2 Belief networks: a concrete model

In real life communication situations, we not only deal with different sources of information with opinions about the facts/events: but also with opinions that these sources may have about each other. We can get information about the weather from a radio broadcasting, a webpage as well as from 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 [15], the revision theoretic framework of Herzberger, Gupta, and Belnap [18,17] suggests a methodology that can be well applied in dealing with these rather complicated situations.

As specified in [15], these situations are represented by directed labelled graphs (DLG). In such graphs, terminal vertices represent facts and non-terminal ones represent agents. Edges, on the other hand, represent agents’ opinions about facts or other agents: 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 . We should mention here that the similar treatment given to the very different concepts of *agent* and *fact* is to keep the model as simple as possible. Nevertheless, there is a difference between them: agents are represented by non-terminal nodes while facts by terminal ones. We assume that if an agent is involved in the situation, she must have some opinion about something or someone (in other words, agents with no opinions are not represented in the model).

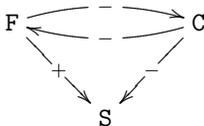
An *external observer* reasons about the communication situations represented by the DLG. While the agents' opinions are represented by edges in the graph, the observer's beliefs are represented in the following way. Vertices are given values from a non-empty finite set A to indicate the states of belief of the observer regarding those agents and facts. As mentioned before, this is a departure from the models in [15], where the value set is a continuous interval rather than 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 (discussed in § 1) leads from an analysis of truth as done in [17,18] to an analysis of a belief network. Consider the following example given in [15].

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



"As in the *Nested Liars*, there are two consistent truth value assignments, but the context makes sure that one of them is intuitively preferred, as the observer's friend has first hand experience of the weather in the street café" ([15]).

Based on this preference, the observer experiences beliefs *flowing* through the graph: the contextually based stronger belief in the node F leads her to believe in S, but at the same time to disbelieve in node C, since it is in conflict with F. Her disbelief in C in turn makes her belief in S much stronger which influences her belief in F once again. Both *forward* and *backward* propagation of beliefs are encountered.

This example shows that in this context of agents and their beliefs, together with the idea of *backward propagation of beliefs* (equivalent to the *backward*

propagation of truth of the pointer semantics [11,12]), there is also *forward propagation of beliefs*. Suppose a trusted source has some positive opinion about a certain proposition φ . Then, 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

We now move on to the formal definition of the model with which we represent the earlier mentioned communication situations.

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 $\ell : \mathcal{E} \rightarrow \{+, -\}$ the labelling function.
- $\Psi : \Lambda^{\mathcal{V}} \rightarrow \Lambda^{\mathcal{V}}$ is the *iteration function*, with Λ the **set of values**.

As mentioned before, vertices in \mathcal{G} represent agents and facts of the communication situation, while edges represent the agents' positive/negative opinions.

The observer's beliefs about agents and facts are represented in a different way. We assume the existence of a function $H : \mathcal{V} \rightarrow \Lambda$, called an *hypothesis*, which assigns 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 .

We want to combine the idea of revision semantics (backward propagation) with forward propagation of beliefs along edges; this is when the iteration function Ψ comes into play. It is used to define a **revision sequence** of belief of the observer in the spirit of revision theory. Given an initial hypothesis H , we define the sequence of functions $\langle H_i; i \in \omega \rangle$ as follows:

$$\begin{aligned} H_0 &:= H \\ H_{i+1} &:= \Psi(H_i). \end{aligned}$$

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:

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

Following Theorem 1 and Theorem 2 in the section 2 of [15], 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

Suppose the observer initially believes in an agent a , but does not have any opinion about agent b . It may well happen that when she comes to know about the positive opinion a has over b , her mental state about having no opinion about b will change to believe in b . The whole idea of the iteration function is to capture the dynamics of the mind-states which cannot be achieved by what we had before. It represents the step-by-step transformation of the observer’s beliefs. With this model, we are able to represent how the observer reasons to finally decide her beliefs about agents and facts in these communication situations.

2.2 Postulates for the iteration function

As mentioned in the introduction, we will consider a finite set A of values for building up the *Assertion Network* model. Without further ado, let us fix such a minimal set that will represent the belief states of the observer. We define the set of values as $A := \{-1, 0, 1\}$, 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. There are many ways in which the iteration function can be defined. Nevertheless, we can ask for it to satisfy some basic properties according to our intuitions of what its behavior should be. Let us first make a brief analysis of what the observer should take into account when deciding her next state of beliefs in facts and agents, given her current ones.

The case of facts is the simple one. In order to get the beliefs the observer has about a fact (represented by a vertex $v \in \mathcal{V}$) at stage $k + 1$ ($H_{k+1}(v)$), the observer should take into account her current beliefs about the fact ($H_k(v)$) and also her current beliefs about the agents that have some opinion (positive or negative) about the fact ($H_k(u)$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$). This is nothing but *forward propagation* of beliefs.

¹ Here, we refer to a propositional language as described in [27]: every propositional variable is an **expression**; \perp and \top are expressions; if E and F are expressions, then $\neg E$, $E \wedge F$, and $E \vee F$ are expressions. If E is an expression and n is a natural number, then $n : E$ is a **clause**.

We say that an **interpretation** is a function $I : \mathbb{N} \rightarrow \{0, 1\}$ assigning truth values to propositional letters. Obviously, an interpretation extends naturally to all expressions. If $n : E$ is a clause and I is an interpretation, we say that I **respects** $n : E$ if $I(n) = I(E)$, and I respects a set of clauses Σ if it respects every element of Σ .

The (partial) stable truth predicate defined by

$$S_H(p_i) := \begin{cases} \text{true} & \text{if } p_i \text{ is stably 'true' in the sequence starting with } H, \\ \text{false} & \text{if } p_i \text{ is stably 'false' in the sequence starting with } H, \text{ and} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

For the details the readers can have a look at [15,27]

The case of an agent is different. Given her current information, how does the observer determine the belief she will have over an agent i (represented by a node $v \in \mathcal{V}$) after one step in her reasoning process? Besides her current beliefs about the agent ($H_k(v)$), she should take into account her current beliefs about the agents that have opinions about i ($H_k(u)$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$; again, *forward propagation*). But not only that; she should also take into account the beliefs she has regarding the 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 put some 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 node of the Assertion Network representing a fact.

1. If at the current stage the observer (a) believes in the agents that have positive opinion about the fact ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$) and (b) does not believe in the agents that have negative opinion about it ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$), then at the next stage she should believe in the fact ($\Psi(H_k(v)) = H_{k+1}(v) = 1$), regardless of her current beliefs about it (the *positive enforcement of facts* postulate).
2. On the other hand, if at the current stage the observer (a) believes in the agents that have negative opinion about the fact ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$) and (b) does not believe in the agents that have positive opinion about it ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$), then at the next stage she should not believe in the fact ($\Psi(H_k(v)) = H_{k+1}(v) = -1$), regardless of her current beliefs about it (the *negative enforcement of facts* postulate).
3. Suppose the observer does not have any opinion about all the agents that have an opinion about the fact ($H_k(u) = 0$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$). Then, the current degree of belief of the observer about the fact should be preserved in the next stage (the *persistence of facts* postulate).

Now let $v \in \mathcal{V}$ be a node representing an agent i_v .

1. If at the current stage the observer (a) believes in the agents that have positive opinion about i_v ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$), (b) does not believe in the agents that have negative opinion about her ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$), (c) believes in the agents/facts about which agent i_v has positive opinion ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle v, u \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$) and (d) does not believe in the agents/facts about which agent i_v has negative opinion ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle v, u \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$), then at the next stage she should believe in agent i_v ($\Psi(H_k(v)) = H_{k+1}(v) = 1$), regardless of her current beliefs about her (the *positive enforcement of agents* postulate).

2. On the other hand, if at the current stage the observer (a) believes in the agents that have negative opinion about i_v ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$), (b) does not believe in the agents that have positive opinion about her ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$), (c) believes in the agents/facts about which agent i_v has negative opinion ($H_k(u) = 1$ for every $u \in \mathcal{V}$ such that $\langle v, u \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "-"$) and (d) does not believe in the agents/facts about which agent i_v has positive opinion ($H_k(u) = -1$ for every $u \in \mathcal{V}$ such that $\langle v, u \rangle \in \mathcal{E}$ with $\ell\langle u, v \rangle = "+"$), then at the next stage she should not believe in agent i_v ($\Psi(H_k(v)) = H_{k+1}(v) = -1$), regardless of her current beliefs about her (the *negative enforcement of agents* postulate).
3. Suppose the observer (a) does not have any opinion about all the agents that have an opinion about i_v ($H_k(u) = 0$ for every $u \in \mathcal{V}$ such that $\langle u, v \rangle \in \mathcal{E}$) and (b) does not have any opinion about all the agents/facts about which i_v has an opinion ($H_k(u) = 0$ for every $u \in \mathcal{V}$ such that $\langle v, u \rangle \in \mathcal{E}$). Then, the current degree of belief of the observer about i_v should be preserved in the next stage (the *persistence of agents* postulate).

2.3 Concrete model

Let us now work towards a concrete definition of the iteration function Ψ . We want to represent backward and forward propagation of beliefs, that is, how the beliefs of the observer about an agent/fact change according to the way they are related to other agents/facts. As discussed before, the observer's new belief about a vertex v should depend on her current beliefs about it and the vertices that can reach v as well as those that can be reached from v .

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

$$\text{In}(v) := \{w \in \mathcal{V}; \langle w, v \rangle \in \mathcal{E}\},$$

that is, $\text{In}(v)$ is the set of vertices that can reach v . We split such set into $\text{In}^+(v)$ and $\text{In}^-(v)$, the sets of vertices that can reach v through a "+"-labelled edge and the sets of vertices that can reach v through an "-"-labelled edge, respectively.

In the same way, we define

$$\text{Out}(v) := \{w \in \mathcal{V}; \langle v, w \rangle \in \mathcal{E}\},$$

that is, $\text{Out}(v)$ is the set of vertices that can be reached from v . We split the set again, so we get $\text{Out}^+(v)$ and $\text{Out}^-(v)$ as above. The set of terminal vertices of \mathcal{G} , denoted by $\mathcal{T}_{\mathcal{G}}$, can be defined as $\mathcal{T}_{\mathcal{G}} := \{v \in \mathcal{V}; \text{Out}(v) = \emptyset\}$.

Now assume the existence of a hypothesis H . For all $w \in \text{In}(v)$, define s_w^v as the H -value of w with sign according to the label of the edge that links v and w , that is,

$$s_w^v := \begin{cases} H(w) & \text{if } w \in \text{In}^+(v) \\ -H(w) & \text{if } w \in \text{In}^-(v) \end{cases}$$

Similarly, for all $w \in \text{Out}(v)$, define t_w^v as the H -value of w with sign according to the label of the edge that links v and w , that is,

$$t_w^v := \begin{cases} H(w) & \text{if } w \in \text{Out}^+(v) \\ -H(w) & \text{if } w \in \text{Out}^-(v) \end{cases}$$

For each value $\lambda \in A$, define S_λ^v as the set of vertices $w \in \text{In}(v)$ such that $s_w^v = \lambda$; similarly, define T_λ^v as the set of vertices in $w \in \text{Out}(v)$ such that $t_w^v = \lambda$.

$$S_\lambda^v := \{w \in \text{In}(v); s_w^v = \lambda\} \quad T_\lambda^v := \{w \in \text{Out}(v); t_w^v = \lambda\}$$

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

$$\Psi(H)(v) := \begin{cases} 1 & \text{if } |S_1^v| > |S_{-1}^v| \\ -1 & \text{if } |S_1^v| < |S_{-1}^v| \\ H(v) & \text{otherwise.} \end{cases}$$

For a non-terminal vertex $v \in \mathcal{V} \setminus \mathcal{T}_G$, 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.

$$IE_v := \begin{cases} 1 & \text{if } |S_1^v| > |S_{-1}^v| \\ -1 & \text{if } |S_1^v| < |S_{-1}^v| \\ H(v) & \text{otherwise.} \end{cases} \quad OE_v := \begin{cases} 1 & \text{if } |T_1^v| > |T_{-1}^v| \\ -1 & \text{if } |T_1^v| < |T_{-1}^v| \\ H(v) & \text{otherwise.} \end{cases}$$

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.*

2.4 Assertion network toolkit

As mentioned in the introduction, looking for the adequate iteration function Ψ is largely an empirical task. At least we can claim that the definition given just now is a plausible one since it satisfies all the intuitive postulates. 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), developed in [15] allows us to play around with the functions and values. It is a piece of graphical interface software

written solely in C/C++ and built on the Boost Graphing Library (BGL), GTK, GTKmm, and the Cairo graphics library. Boost property maps contain all the data of a graph object while the drawing area and table widgets serve as a view for the values stored in the maps. Stepping through a function locks the map and updates the view according to the new values.

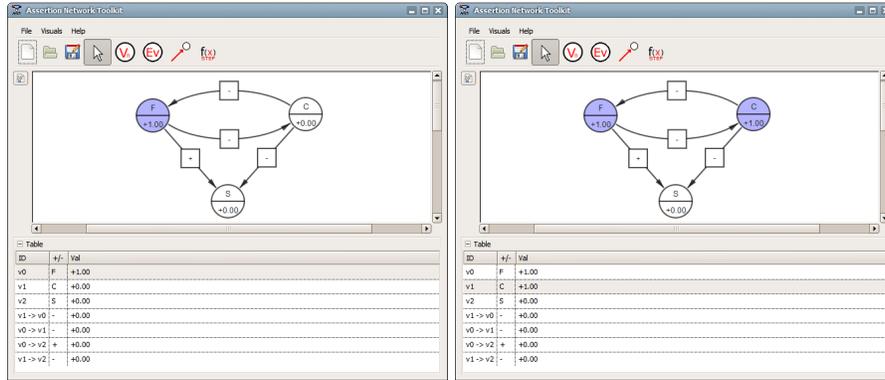


Fig. 1. Initial opinions

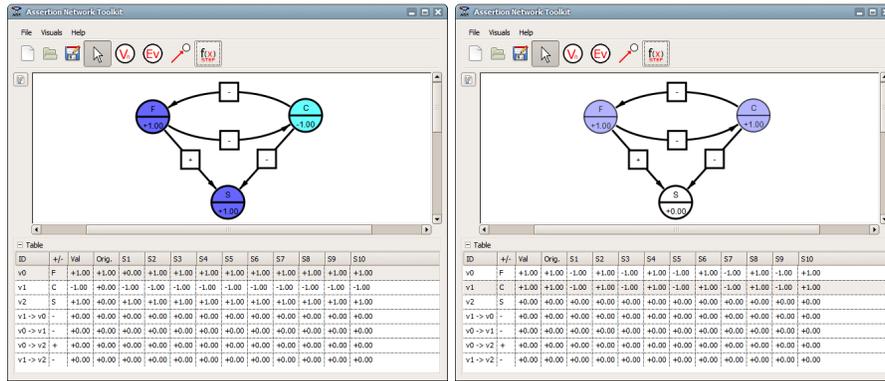


Fig. 2. Final opinions

As an example of its use, consider the communication situation described at the beginning of this section. The iteration function Ψ defined before 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 figure of Figure 1), as mentioned in the example earlier, the observer’s friend has a first hand experience of what is happening outside the office, so the observer is inclined to believe him; 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 ($H_0(\mathbf{F}) = 1$) and a value of 0 to the others ($H_0(\mathbf{C}) = 0$ and $H_0(\mathbf{S}) = 0$). The second one (the right hand side figure of Figure 1), depicts the case that for some reason whatsoever, the observer has an equally high opinion about her friend as well as her colleague in the beginning. Here, the initial hypothesis H_0 assigns a value of 1 to both her friend ($H_0(\mathbf{F}) = 1$) and colleague ($H_0(\mathbf{C}) = 1$), but 0 to the mentioned fact ($H_0(\mathbf{S}) = 0$).

We let the program iterate the function several times. As output of the process (Figure 2), we see that though in the first case all the vertices reach stable values, in the second case only the vertex representing the fact gets a stable value, that of ‘no opinion’ of the observer. The readers will definitely consent to the fact that in both these cases, the final opinion values completely agree with our intuitions.

The sequence of values generated by the iteration (see the tables below) indicates us that in the first case it takes just two steps to reach such stable values, and in the second case the values of \mathbf{F} and \mathbf{C} oscillate.

	H_0	H_1	H_2	H_3	\dots
\mathbf{F}	1	0	1	1	\dots
\mathbf{C}	0	-1	-1	-1	\dots
\mathbf{S}	0	1	1	1	\dots

	H_0	H_1	H_2	H_3	\dots
\mathbf{F}	1	-1	1	-1	\dots
\mathbf{C}	1	-1	1	-1	\dots
\mathbf{S}	0	0	0	0	\dots

3 Expressing belief networks

In this section we provide a logical language with which we can express the behavior of the Assertion Network Model. Since the network focuses on the observer’s point of view, we define a language that takes her perspective. Our atomic propositions are expressions indicating the state of belief the observer has about some agent or fact 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 what is happening in the situation represented by the graph (*the observer believes in agent i in the beginning but may not have any opinion about i anymore, after considering the surroundings*). It will serve our purpose here. Readers who are interested in a language describing the graph can have a look at [16].

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 \circ 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 $\circ\varphi$ are interpreted as “*after one iteration of the function, φ is the case*”. We want this next-step operator since we intend to describe the way the beliefs of the 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 a syntactic operator \oplus ; 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 $\oplus\varphi$ are interpreted as “*after the whole process, φ is the case*”.

3.1 A language expressing the observer’s beliefs

To express the observer’s epistemic attitudes about the communication situations, we define the following language which is extensional in nature. Given a set of *agents* \mathbf{A} , a set of *propositions* \mathbf{P} , the *Language of Beliefs* \mathcal{LB} is given by:

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

with $\gamma \in \mathbf{A} \cup \mathbf{P}$. Formulas of the form $\mathbf{B}\gamma$ indicates “*the observer believes in γ* ”, while $\mathbf{N}\gamma$ indicates “*the observer does not have any opinion about γ* ” and $\mathbf{D}\gamma$ 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. Though we may have simply avoided \mathbf{B} , \mathbf{N} and \mathbf{D} from the atomic propositions, they are incorporated to make the language more intuitive, keeping in mind the formal interpretations of these formulas, as defined later.
- The language \mathcal{LB} is *not a modal language*. Its atomic propositions $\mathbf{B}\gamma$, $\mathbf{N}\gamma$ and $\mathbf{D}\gamma$ have special meanings, but they are still atomic propositions.
- Propositional languages evaluate formulas by assigning truth values to basic propositions and then extending it to more complex formulas. Usually, there is no relation between the atomic propositions, so the truth value of one of them is independent of the truth value of the others. In our language, the semantics of atomic propositions will be defined in a way such that the truth value of some of them is related: formulas like $\mathbf{B}\gamma \wedge \mathbf{D}\gamma$, for example, will never be true.

Consider an Assertion Network Model $M = (\mathcal{G}, \Psi)$ with $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \ell)$. In order to use it to give meaning to formulas in \mathcal{LB} , we need to uniquely identify each vertex in \mathcal{V} with an agent or a fact in $\mathbf{A} \cup \mathbf{P}$. Moreover, the map should satisfy the already mentioned requirement: facts should be mapped to terminal vertices. An *interpretation* I is a partial injective function $I : \mathbf{A} \cup \mathbf{P} \rightarrow \mathcal{V}$ such that, for each $p \in \mathbf{P}$, we have $I(p) \in \mathcal{T}_{\mathcal{G}}$, when it is defined.

Given an interpretation I and an initial hypothesis H , the truth definition of formulas of \mathcal{LB} in an Assertion Network Model $M = (\mathcal{G}, \Psi)$ is given by

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

Thus, the formula $\mathbf{B}\gamma$ 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; similarly $\mathbf{N}\gamma$ corresponds to the value 0 and $\mathbf{D}\gamma$ to -1.

3.2 A language expressing belief flow

The language \mathcal{LB} allows us to express just the observer's beliefs about a communication situation, but does not permit us to express how the opinions change as a result of the belief propagation. Our first step towards such a goal is to extend the language to express what happens after one step in the observer's introspection process. For that purpose, the language is extended by closing it under the \circ operator: if φ is a formula of the language, so it is $\circ\varphi$.

To give truth value to the new formulas, we use the iteration function; this way we can modify the hypothesis H . Formally, we have:

$$M, I, H \models \circ\varphi \quad \text{iff} \quad M, I, \Psi(H) \models \varphi$$

Now we have formulas like $\mathbf{B}\gamma \wedge \circ\mathbf{N}\gamma$, expressing “*originally, the observer believes in γ , but after one introspection step (that is, after considering once all the information she has), she does not have an opinion anymore.*”

Finally, to get to the logic of *Assertion Network Semantics* proposed in the earlier section, we introduce the operator \otimes in the language to talk about the stable positions in the model. The full language of the *Logic of Belief Flow* \mathcal{LBF} , is given by:

$$\varphi := \mathbf{B}\gamma \mid \mathbf{N}\gamma \mid \mathbf{D}\gamma \mid \neg\varphi \mid \varphi \vee \psi \mid \circ\varphi \mid \otimes\varphi$$

with $\gamma \in \mathbf{A} \cup \mathbf{P}$, as before.

Formulas of the form $\circ\varphi$ express “*after the observer considers once all the information she has, φ is the case*”. The new operator \otimes is used to express the observer's state of beliefs after the whole introspection process: formulas of the form $\otimes\varphi$ express “*after the observer considers all the information she has, φ is the case*”. Its truth definition is given by:

$$M, I, H \models \otimes\varphi \quad \text{iff} \quad \exists n \in \omega \text{ such that } M, I, \Psi^i(H) \models \varphi \text{ for all } i \geq n.$$

In the new language we have formulas like $\mathbf{B}\gamma \rightarrow \otimes\mathbf{D}\gamma$, expressing “*originally, the observer believes in γ , but after the introspection process (that is, after considering all the information she has), she does not believe in γ anymore*”.

Now we have a language with which we can express the observer's initial beliefs, and also how such beliefs change as a result of the introspection process,

either in one step (with the operator \circ) or after the whole process (with the operator \otimes). 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 in page 11.

Formally, the Assertion Network Model and the interpretation are given together by $M = ((\mathcal{V}, \mathcal{E}, \ell), \Psi)$ with

- $\mathcal{V} := \{\mathbf{F}, \mathbf{C}, \mathbf{S}\}$,
- $\mathcal{E} := \{\langle \mathbf{F}, \mathbf{C} \rangle, \langle \mathbf{C}, \mathbf{F} \rangle, \langle \mathbf{F}, \mathbf{S} \rangle, \langle \mathbf{C}, \mathbf{S} \rangle\}$,
- $\ell\langle \mathbf{F}, \mathbf{S} \rangle = \text{“+”}$; $\ell\langle \mathbf{F}, \mathbf{C} \rangle = \ell\langle \mathbf{C}, \mathbf{F} \rangle = \ell\langle \mathbf{C}, \mathbf{S} \rangle = \text{“–”}$.

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

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

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

- $\mathbf{BF} \wedge \mathbf{NC} \wedge \mathbf{NS}$
- $\circ(\mathbf{NF} \wedge \mathbf{DC} \wedge \mathbf{BS})$
- $\circ \circ (\mathbf{BF} \wedge \mathbf{DC} \wedge \mathbf{BS})$
- $\otimes(\mathbf{BF} \wedge \mathbf{DC} \wedge \mathbf{BS})$

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

- $(\mathbf{BF} \wedge \mathbf{BS}) \rightarrow \otimes(\mathbf{BF} \wedge \mathbf{DC} \wedge \mathbf{BS})$
If the observer initially believes in F and S , then her initial belief about C is irrelevant.
- $(\mathbf{BF} \wedge \mathbf{BC} \wedge \mathbf{NS}) \rightarrow ((\circ^k \mathbf{BF} \rightarrow \circ^{k+1} \mathbf{DF}) \wedge (\circ^k \mathbf{DF} \rightarrow \circ^{k+1} \mathbf{BF})) \quad (k \geq 0)$
If she initially believes in F and C without having an opinion about S , then her beliefs about F will oscillate ($\circ^0 \varphi := \varphi$ and $\circ^{k+1} \varphi := \circ \circ^k \varphi$).
- $(\mathbf{BF} \wedge \mathbf{BC} \wedge \mathbf{NS}) \rightarrow \neg \otimes (\mathbf{BF} \vee \mathbf{NF} \vee \mathbf{DF})$
Therefore, there is no stable value for F .
- $(\mathbf{BF} \wedge \mathbf{BC} \wedge \mathbf{NS}) \rightarrow \otimes \mathbf{NS}$
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:

- $\mathbf{B}\gamma \rightarrow (\neg \mathbf{N}\gamma \wedge \neg \mathbf{D}\gamma)$
- $\mathbf{D}\gamma \rightarrow (\neg \mathbf{N}\gamma \wedge \neg \mathbf{B}\gamma)$
- $\mathbf{N}\gamma \rightarrow (\neg \mathbf{B}\gamma \wedge \neg \mathbf{D}\gamma)$
- $\circ(\varphi \wedge \psi) \leftrightarrow (\circ\varphi \wedge \circ\psi)$
- $\otimes(\varphi \wedge \psi) \leftrightarrow (\otimes\varphi \wedge \otimes\psi)$
- $\otimes\varphi \rightarrow \otimes \circ \varphi$

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 in a multi-agent setting, and it is still going on. In the following we provide a discussion so as to highlight the novelty of our approach in the current setting both from the modelling as well as the logical perspectives.

4.1 Different approaches for revising/merging

As mentioned, there is a huge amount of existing literature in area of revising and/or merging of beliefs. Some of them have a logical perspective, providing postulates, models and formal languages to describe the respective processes. Some others have a connectionist perspective, considering the phenomena as the emergent processes of interconnected networks of simple units (usually, neural network models).

The classical work on belief revision is the so called *AGM approach* [1], where the authors presented postulates that an operator which performs revision should satisfy in order to be considered rational. Several other frameworks for modelling revision have been proposed. Particularly related with our proposal are those focused on *iterated* revision, like [5] and [19]. 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 ([20,21,22]).

Our approach lies on the revision side, with the agents' opinions and the observer's beliefs are 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 [13]. 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 for modelling such information change, like dynamic epistemic logic (*DEL*; [14,28]) and dynamic doxastic logic (*DDL*; [29]). In [30] the author looks into *DEL* together with the *AGM* style belief revision theory [1] and provides 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, *LB \mathcal{F}* focusses on introspection of a reasoning agent regarding the transition of her belief states in a communication situation. While the belief states can be easily expressible in a propositional language, their transition is captured by the dynamic modal operators \circ and \otimes . It should be mentioned here that the underlying belief flow of the observer regarding the agents and the events taking part in the interaction gets portrayed in a very illustrative way. 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 capable of changing one's knowledge state, as well as *soft* information by which her beliefs get affected. On the other hand, *LB \mathcal{F}* 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, we have approaches that are based on interconnected networks, where the result of the process is related with the values that become stable once the information has propagated along the net.

In [25,26], a *Neural-Logic Belief Network (NLBN)* is defined, which is basically a hybrid symbolic-neural network structure - a finite acyclic directed weighted graph with nodes representing propositions and edges representing relations between these propositions. Different categories of nodes, viz. input nodes and base nodes as well as that of edges like combinative and alternative links are considered, together with sets of deterministic operations for node value propagation and also for dynamic belief changes. The network propagation functions consist of neural-net feed-forward computations together with degree-of-belief computations which are symbolic in nature and they provide a stable belief state in any *NLBN*, where belief states are finite sets of propositions providing positive or negative information, together with their associated ordinal belief-values. As exemplified in the above mentioned papers, *NLBN* can be used to model common sense reasoning, and provides a way for representing transitions in the belief attitudes of agents.

A very different approach to belief revision has been provided in [8], viz. a distributed approach. Models have been proposed to represent the collective activity of a group of interacting agents, taking into consideration the local beliefs of the individual components and their communication with each other. The assimilation of the different beliefs is done by the whole group through an election mechanism. Several simulation experiment were performed so as to compare between the group action under a distributed architecture, where the agents can communicate with each other and a more centralized architecture where it is the task of a single agent to combine the different opinions of the group. Numerical values as probability measures have been incorporated in these models to represent degrees of uncertainty, and computation is performed using Dempster rule and Bayesian conditioning.

The discussion here will not be complete if we do not mention Bayesian belief nets (*BBN*) [31]. They are really powerful tools modelling cause and effect relationship in a wide range of fields. *BBNs* are basically directed acyclic graphs, with nodes representing the variables and the edges representing their causal relationship. Each node (variable) corresponds to a set of its states (probable values). The probability that a node is in one of its corresponding states is given by a conditional probability table, based on the causal relationship. Based on these tables, *BBN* can be used in decision making, where both inductive as well as deductive reasoning can be performed, the basic computations being done with respect to Bayes' rule.

Let us now focus on our model proposed in §2 in this perspective. Real life communication situations talking about facts/events, agents and their opinions are represented by directed cyclic labelled graphs. Both forward and backward propagation of belief flow are taken into consideration. The computation function is a simple three-valued one, which is deterministic, also. The novelty lies in the semantics derived from stability as used in the revision theory of truth [18,17]. In *NLBN*, only forward propagation of belief is considered and the representation is restricted to propositions only. Similar is the case in *BBN*, though in some sense, *BBN*, with its probabilistic computation approach can be used in a greater variety of domains. The work of Dragoni and Giorgini [8] is closer to our approach though with subtle but important differences, as the readers can easily notice. The most notable among them is that our approach is very centralized in nature.

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. The work we mention now is a nice combination of the two perspectives.

In [13], the authors propose a conciliation operation via iterated belief merging. The beliefs of several agents are first merged to create a new belief set, and then each agent revises her own initial beliefs with respect to the result of the merging. The process is iteratively repeated until a fixed point is reached, and the conciliation operator is defined with respect to it. The paper explores two strategies for the revision steps, relative to the agent’s confidence in her own beliefs: credulous, in which the result of the merging takes precedence over the agent’s beliefs, or sceptical, where the agent gives priority to what she believes.

This work is similar to ours in the sense that it also looks for a *stable* situation where further interaction between the diverse components will not modify the current status of the model. Unlike our approach, they use a two-stage iterative process: the first one merging the beliefs of all the different agents, and the second one allowing each agent to revise her own initial beliefs in accord with the merging process. 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 also not new. In fact, it has been a proposed solution for the so called *logical omniscience* problem. This problem is about unrealistic assumptions on the reasoning power of the agents. In epistemic logic with Kripke semantics, for example, the knowledge of the represented agents includes all validities and is closed under logical consequence.

In [9,10], 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, but also to allow her to improve her information set as time goes by. Agents represented within this

framework are not logically omniscient, since the sets of formulas do not have to contain all validities and do not have to be closed under logical consequence. Nevertheless, they are not logical ignorant also: at every point of time their information gets increased, and eventually they will get to know all validities and the consequences of their initial knowledge. Instead of representing agents that know everything (they are supposed to know) from the very beginning, this approach focusses on the step-by-step process that leads to that outcome.

Our proposed language \mathcal{LBF} shares this concept. Instead of considering the revision process as a single-step operation, we focus on the small steps that may or may not lead to an outcome. In some situations, the revision process will in fact lead to a sequence of stable values, indicating that the (possible inconsistent) initial information and the observer’s initial attitudes can fit together. In some others, like with the liar sentence, the beliefs of the observer oscillate, indicating that the initial information and the observer’s initial attitudes 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 the opinion the observer has about the discussed facts, but also the opinion she has about other agents. These opinions basically indicate if the observer *trusts* the agents or not. Several efforts have been made to analyze the notion of *trust* in multi-agent systems. The notion is important since it allows us to represent asymmetries in the way the agents’ opinions at some stage will influence the observer’s beliefs at the next one.

In [24], Liau proposes a modal logic with three modal operators: B_i (a normal KD45 operator), I_{ij} (a normal KD operator) and T_{ij} (a non-normal modal operator). Formulas of the form $B_i\varphi$ express agent i ’s beliefs, $I_{ij}\varphi$ ’s represent communication indicating that agent i acquires information φ from agent j , and T_{ij} indicate that agent i trust the judgment of agent j about φ .

Since beliefs and information are normal modal operators, an agent’s belief set is 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: if an agent i trust in the judgement of j about φ , she does not have to trust j about the judgment of all logical consequences of φ .

In [6], 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 a question (*question-based trust derivation* and *question-based distrust derivation*). Consider an agent j answering positively to a question of agent i about φ : if i believes in φ , she will trust j about φ , if i believes in $\neg\varphi$, she will *not* trust j about φ . This is somewhat similar to the *backward propagation* concept.

Let us now come back to our proposed framework. It is to be noted that the notion of belief in an agent, analogous to the notion of trust of the described works, is not relative to a particular statement of the agent (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 .

As mentioned, our approach does not ask for particular initial properties of the observer’s beliefs and the agents’ opinions. In most of the revision approaches, the initial beliefs should be consistent, but by removing such restriction we allow a more general initial setting. Our observer’s reasoning does not start from an ideal state, but it will reach it whenever possible.

The just described works consider properties of the relations between beliefs, information and trust, but they work on a static approach without considering *dynamics* of the system. Even exchanges of information and questions are semantically represented as properties of the model (properties on the corresponding binary relations) and not as actions that modifies it. On the other hand, the main focus of our approach is the *dynamic* process through which all the involved participants start working, updating the model and influencing themselves while trying to reach an agreement (a stable consistent situation).

5 Conclusion and intentions

In this work, we propose a model of belief propagation on a graph representing communication situations 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 also compared this model as well as the language developed with some of the existing frameworks for modelling communication situations.

As we have mentioned, the belief about a particular node in the graphical model at the next stage is given in terms of the beliefs at this stage about the incoming and outgoing nodes (forward and backward propagation) corresponding to the concerned node. The postulates we have stated leave open the way the iteration function should behave in more general situations. In situations where the balance is not completely biased to one side, there is no unique way for the observer of determining her beliefs at the next stage.

Different kinds of observers can be represented by iteration functions with different properties. A *majority* function may represent an observer that follows what the majority says; a *confident* function can represent an agent that gives more weight to her current beliefs while a *credulous* one can represent observers that give more precedence to others’ opinions. We can even consider a *preferential* function, representing observers that give precedence to agents for which she has positive beliefs. It will be interesting to formalize those different policies. Some further possible avenues for future work are as follows.

Opinionated edges One way of refining this approach is to consider beliefs about agents' opinions also. We can think of situations where a given agent is an expert in a given subject but not in some other. In our current model, the observer has to decide the degree of belief over the agent, regardless of the facts the agent has opinions about, but in some cases it is more natural to have different degrees of beliefs in the agent's different opinions. Semantically, this would extend the domain of the hypothesis to the set $\mathcal{V} \cup \mathcal{E}$ and will force a new definition of the iteration function Ψ .

Extending value set The degrees of belief of the observer range over the members of the value set Λ . Here we have considered a three-valued Λ , but it can easily be extended to any finite valued one, so as to express more possible epistemic states of an agent. Things will get more complicated this way but closer to the actual real life situations. This will also involve a redefinition of the iteration function Ψ .

Comparing expressivity On the logical side, the presented language allows us to express the observer's current beliefs, how they change after steps in the introspection process and whether any stability is attained. Formulas of the form $\otimes\varphi$ expressing stable values also relates 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

The authors gratefully acknowledge Erik Scorelle for working on *Assertion Network Toolkit* with the modified definitions and helping in with the screenshots. They 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.

References

1. C. Alchourrón, P. Gardenfors, and D. Makinson. On the logic of theory change. *The Journal of Symbolic Logic*, 50:510–530, 1985.
2. Christian Balkenius and Peter Gärdenfors. Nonmonotonic inferences in neural networks. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Principles of knowledge representation and reasoning*, pages 32–39, San mateo, CA, 1991. Morgan Kaufmann.
3. Reinhard Blutner and Paul David Doherty. Nonmonotonic logic and neural networks. Available at <http://citeseer.ist.psu.edu/160749.html>.
4. Thomas Bolander. Self-reference and logic. *Phi News*, 1:9–44, 2002.
5. Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. In Ronald Fagin, editor, *Proceedings of the Fifth conference on Theoretical aspects of reasoning about knowledge (TARK '94)*, pages 5–23, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

6. Mehdi Dastani, Andreas Herzig, Joris Hulstijn, and Leendert W. N. van der Torre. Inferring trust. In João Alexandre Leite and Paolo Torroni, editors, *CLIMA V*, volume 3487 of *Lecture Notes in Computer Science*, pages 144–160. Springer, 2004.
7. Artur d’Avila Garcez, Krysia Broda, and Dov Gabbay. Symbolic knowledge extraction from trained neural networks: A sound approach. *Artificial Intelligence*, 125(1–2):155–207, January 2001.
8. A.F. Dragoni and P. Giorgini. Distributed belief revision. *Autonomous Agents and Multi-Agent Systems*, 6(2):115–143, 2003.
9. Ho Ngoc Duc. Logical omniscience vs. logical ignorance on a dilemma of epistemic logic. In *EPIA ’95: Proceedings of the 7th Portuguese Conference on Artificial Intelligence*, pages 237–248, London, UK, 1995. Springer-Verlag.
10. Ho Ngoc Duc. Reasoning about rational, but not logically omniscient, agents. *Journal of Logic and Computation*, 7(5):633–648, 1997.
11. Haim Gaifman. Operational pointer semantics: Solution to self-referential puzzles i. In Moshe Vardi, editor, *Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 43–59, 1988.
12. Haim Gaifman. Pointers to truth. *Journal of Philosophy*, 89:223–261, 1992.
13. Olivier Gauwin, Sébastien Konieczny, and Pierre Marquis. Conciliation through iterated belief merging. *Journal of Logic and Computation*, 17(5):909–937, October 2007.
14. Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation (University of Amsterdam), 1999.
15. Sujata Ghosh, Benedikt Löwe, and Erik Scorelle. Belief flow in assertion networks. In Uta Priss, Simon Polovina, and Richard Hill, editors, *Proceedings of the 15th International Conference on Conceptual Structures (ICCS 2007)*, Sheffield, UK, volume 4604 of *LNAI*, pages 401–414, Heilderberg, July 2007. Springer-Verlag.
16. Sujata Ghosh and Fernando R. Velázquez-Quesada. Merging information.
17. A. Gupta and Nuel D. Belnap. *The Revision Theory of Truth*. The MIT Press, Cambridge, MA, 1993.
18. Hans G. Herzberger. Naive semantics and the liar paradox. *The Journal of Philosophy*, 79:479–497, 1982.
19. Yi Jin and Michael Thielscher. Iterated belief revision, revised. *Artificial Intelligence*, 171(1):1–18, 2007.
20. Sébastien Konieczny. *Sur la logique du changement: révision et fusion de bases de connaissance*. PhD thesis, Laboratoire d’Informatique Fondamentale de Lille, Université de Lille 1, November 1999.
21. Sébastien Konieczny and Ramón Pino-Pérez. On the logic of merging. In *Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR ’98)*, pages 488–498, 1998.
22. Sébastien Konieczny and Ramón Pino-Pérez. Merging information under constraints: a logical framework. *Journal of Logic and Computation*, 12(5):773–808, 2002.
23. Hanner Leitgeb. Interpreted dynamical systems and qualitative laws: from neural networks to evolutionary systems. *Synthese*, 146(1-2):189–202, August 2005.
24. Churn-Jung Liau. Belief, information acquisition, and trust in multi-agent systems: a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.
25. B.T. Low and N.Y. Foo. Towards a network for representing beliefs. In *Proceedings of the 2nd International Computer Science Conference*, pages 85–91, 1992.
26. B.T. Low and N.Y. Foo. A network formalism for commonsense reasoning. In *Proceedings of the Sixteenth Australian Computer Science Conference*, pages 425–434, 1993.

27. Benedikt Löwe. Revision forever! In Henrik Schärfe, Pascal Hitzler, and Peter Øhrstrøm, editors, *Proceedings of the 14th International Conference on Conceptual Structures (ICCS 2006), Aalborg, Denmark*, volume 4068 of *LNAI*, pages 22–36, Berlin, July 2006. Springer-Verlag.
28. J. A. Plaza. Logics of public communications. In M. L. Emrich, M. S. Pfeifer, M. Hadzikadic, and Z. W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.
29. K. Segerberg. The basic dynamic doxastic logic of AGM. In M.A. Williams and H. Rott, editors, *Frontiers in Belief Revision*, pages 57–84. Kluwer, Dordrecht, 2001.
30. J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, 17(2):129–155, 2007.
31. Jon Williamson. *Bayesian nets and causality, Philosophical and computational foundations*. OUP, 2005.