# How strong to believe information ? a trust model in the logical belief function theory

Laurence Cholvy. ONERA 2 avenue Ed. Belin 31055 Toulouse FRANCE laurence.cholvy@onera.fr

### Abstract

To which extent an agent can believe a piece of information it gets? This is the question we address in this paper. More precisely, we provide a model for expressing the relations between the trust an agent puts in the information sources and its beliefs about the information they provide. This model is based on Demolombe's model and extends it by considering imperfect information sources i.e information sources that can report false information. Furthermore, this model is defined in the belief function theory allowing degrees of trust to be modelled in a quantitative way. Not only this model can be used when the agent directly gets information from the source but it can also be used when the agent gets second hand information i.e., when the agent is not directly in contact with the source.

## 1 Motivation

When it gets a new piece of information from an information source, a rational agent has to update or revise its belief base if it considers this piece of information sufficiently supported [1], [20], [12], [13]. Thus an important question for the agent is to estimate how this new piece of information is supported i.e, how strong it can believe it ?

Obviously, the agent may believe a new piece of information if it trusts the information source for delivering true information [7], [14]. For instance, assume that in order to know if it will rain tomorrow, I look at Météo-France web site and read that indeed it will rain. If I trust Météo-France for delivering correct forecast, then I can believe what Météo-France is reporting i.e., I can believe that it will rain.

But in many applications [19], [16] [17], information sources are not necessarily correct all the time. They may deliver false information, intentionnally or not. This is why, assuming them to be incorrect must also be useful. Furthermore, it may happen that information of interest is second hand i.e., the agent who collects information is not directly in contact with the information source: it obtains information through an agent which cites the information source. The process can even be longer. This is the case when, for instance, I am informed by my neighbour that he read on Météo-France web site that it is going to rain. In this case, the information my neighbour reports is that he read that according to Météo-France it is going to rain. My neighbour does not

Copyright © by the paper's authors. Copying permitted only for private and academic purposes.

In: R. Cohen, R. Falcone and T. J. Norman (eds.): Proceedings of the 17th International Workshop on Trust in Agent Societies, Paris, France, 06-MAY-2014, published at http://ceur-ws.org

tell me that it is going to rain. We insist on the fact that here, the very information which interests me (will it rain ?) is reported via two agents, Météo-France and my neighbour, the second citing the first. Second hand information management is a more delicate issue. Indeed, trusting my neighbour for giving true forecast is not useful here. However, trusting him not to lie and trusting Météo France for giving true forecast will lead me to believe that it is going to rain.

One very interesting contribution in the trust community is Demolombe's one [7],[8],[9]. The author considers trust as an attitude of an agent who believes that another agent has a given property. As for an agent which provides information (i.e an information source), it can be attributed six properties: it can be sincere (roughly speaking, it believes what it reports), competent (information it believes is true), valid (information it reports is true), cooperative (it reports what it believes), vigilant (it believes what is true) or complete (it reports what is true). In this model, validity is defined as sincerity and competence and completeness is defined as cooperativity and vigilance. This work shows that the trust an agent puts in an information source relatively to some of these properties influences the fact that this agent believes what this source produces or does not produce. This is formally shown by considering a formal logical framework expressed in modal logic. The operators of this modal logic are :  $B_i$  ( $B_i p$  means "agent *i* believes that p") and  $I_i^j$  ( $I_i^j p$  means "agent *i* informs<sup>1</sup> agent *j* that p"). Operators  $B_i$  obey KD system and their semantics are defined by serial accessibility relations between words. Operators  $I_i^j$  satisfy  $I_i^j p \wedge I_i^j q \to I_i^j (p \wedge q)$  and obey rule of equivalence substitutivity [2]. Their semantics are defined by neighborhood functions. Furthermore, it is assumed that there is no failure in the communication process when an agent informs another one. I.e,  $I_i^j p \to B_i I_i^j p \to B_i \neg I_i^j p$  are considered as true. The different sorts of trusts considered in this model are<sup>2</sup>:

- $Tsincere_{a,b}(p) =_{def} B_a(I_b^a p \to B_b p)$  i.e., a trusts b for p in regard to sincerity iff a believes that if b tells it p then b believes p.
- $Tcompetent_{a,b}(p) =_{def} B_a(B_b p \to p)$  i.e., a trusts b for p in regard to competence iff a believes that if b believes p then p is true.
- $Tcooperative_{a,b}(p) =_{def} B_a(B_p p \to I_b^a p)$  i.e., a trusts b for p in regard to cooperativity iff a believes that if b believes p then b tells it p.
- $Tvigilant_{a,b}(p) =_{def} B_a(p \to B_b p)$  i.e., a trusts b for p in regard to vigilance iff a believes that if p is true then b believes it.
- $Tvalid_{a,b}(p) =_{def} B_a(I_b^a p \to p)$  i.e., a trusts b for p in regard to validity iff a believes that if b tells it p then p is true.
- $Tcomplete_{a,b}(p) =_{def} B_a(p \to I_b^a p)$  i.e., a trusts b for p in regard to completeness iff a believes that if p is true then b tells it.

These notions can then formally be used to derive the beliefs of an agent who receives a piece of information. For instance,  $Tvalid_{i,j}(p) \to (I_j^i p \to B_i p)$  is a theorem. It shows that if agent *i* trusts agent *j* for information *p* in regard to validity, then if *j* tells it *p* then *i* believes *p*. An instance of this theorem is:  $Tvalid_{i,MF}(rain) \to (I_{MF}^i rain \to B_i rain)$  which means that if I trust Météo-France for being valid for forecast, then if Météo-France reports that it will rain then I believe that it will rain. Notice that we also have the following theorem:  $Tcomplete_{i,MF}(rain) \to (\neg I_{MF}^i rain \to B_i \neg rain)$ . This means that if I trust Météo-France for being complete for forecast, then if Météo-France does not report that it will rain then I believe that it will rain. Notice that we also have the following theorem:  $Tcomplete_{i,MF}(rain) \to (\neg I_{MF}^i rain \to B_i \neg rain)$ . This means that if I trust Météo-France for being complete for forecast, then if Météo-France does not report that it will rain then I believe that it will not rain.

This model can be used to make more complex inferences, in particular for the problem of trust propagation when agents inform each other about their own trusts [9]. For instance, suppose that due to a gaz failure at home, my husband calls a plumber. Suppose that my husband tells me that the plumber told him that our gaz installation must be renewed. He adds that he trusts this person because he owns the Qualigaz agreement. Then, if I trust my husband to tell the truth, I can conclude that our gaz installation must be renewed.

Furthermore, agents may be uncertain about the trust they put in information sources. In [8], Demolombe presented a qualitative framework to model this uncertainty and defines a modal logic for graded beliefs. Graded beliefs are modelled by operators  $B_i^g$  so that  $B_i^g p$  means that the strength level of agent *i* about *p* is *g*. However,

<sup>&</sup>lt;sup>1</sup>We will sometimes say "tells" or "reports" or "produces"

<sup>&</sup>lt;sup>2</sup>We adopt here the definitions given in [9] which defines trust with the belief operator  $B_i$  and not the definitions of [7] which defines trust with a "strong belief"  $K_i$  operator

the author did not show the application of this framework to second hand information management. In [6], Pichon et al address very close questions by using the belief function theory [18] which is a framework for modelling uncertainty in a quantitative way. They propose a mechanism for computing the plausibility of a piece of information which is emitted by an agent given our uncertain belief about its reliability. For the authors, the reliability of an agent is defined by its relevance and its truthfulness so that (1) information provided by a non-relevant information source is ignored i.e a non-relevant source brings no information; (2) we can believe the piece of information provided by a relevant and truthful source; (3) we can believe the negation of the piece of information provided by a relevant but non-truthful source. It must be noticed that "being relevant and truthful" is quite close to "being valid" as introduced by Demolombe. However, this work did not address the case of second hand information either.

In the same time, we suggested to extend Demolombe's model by considering the negative counterparts of the previous six properties [3], [4]. In particular, we considered negative counterparts of validity and completeness, which led us to consider misinformer agents and falsifier agents. Roughly speaking, when a misinformer reports a piece of information, then we can conclude that it is false. When a piece of information is false, a misinformer reports it is true. In Demolombe's terms, this could be defined by:

- $Tmisinformer_{a,b}(p) =_{def} B_a(I_b^a p \to \neg p)$  i.e., a trusts b for being misinformer relatively to p iff a believes that if b tells it p then p is false. Notice that "being misinformer" is quite close to "being relevant and non truthful" as introduced by Dubois et al.
- $Tfalsifier_{a,b}(p) =_{def} B_a(\neg p \rightarrow I_b^a p)$  i.e., a trusts b for being falsifier relatively to p iff a believes that if p is false then b tells it is true.

In the present paper, we go on our research by proposing a model which (i) allows an agent to express to which extent it trusts an information source for being valid, complete, misinformer or falsifier; (ii) which can also be used to manage second hand information. This model is defined in a framework recently defined, the logical belief function theory [5]. This framework extends the belief function theory and allows the user to assign degrees of beliefs to propositional formulas and to express integrity constraints.

This paper is organized as follows. Section 2 describes a model which allows an agent to express uncertainty about the trust it puts in sources. Section 3 shows how to use this model to evaluate information. We first study the case when the agent is in direct contact with the source and then the case when it is not. Examples will illustrate the model. Finally section 4 is devoted to a discussion.

## 2 A trust model in the logical belief function theory

The belief function theory [18] is a framework which offers several interesting tools to manage uncertain beliefs such as mass functions and belief functions which are used to quantify the extent to which one can believe a piece of information and rules of combination which are used to combine uncertain beliefs in many different way [10]. Ignorance can explicitly be quantified in this framework and consequently this model does satisfy the rule of additivity, i.e., the degree of belief in a proposition plus the degree of belief in its negation can be strictly less than 1.

In this paper, we will use the *logical belief function theory* [5] which extends the belief function theory by allowing the user to express its beliefs in propositional logic and to consider integrity constraints. It has been proved that this formalism is not more expressive than the belief function theory. However, it allows to express uncertain beliefs directly and in a more compact way. Furthermore, it enlightens the role of integrity constraints. This framework is summarized below.

#### 2.1 The logical belief function theory

Let  $\Theta$  be a finite propositional language and  $\sigma$  be a satisfiable formula of  $\Theta$ .

We consider the equivalence relation denoted  $\stackrel{\sigma}{\rightarrow}$  which is defined by:  $A \stackrel{\sigma}{\rightarrow} B$  iff  $\sigma \models A \leftrightarrow B$ . Thus, two formulas are in relation by  $\stackrel{\sigma}{\rightarrow}$  if and only if they are equivalent when  $\sigma$  is true. By convention, we will say that all the formulas of a given equivalent class of  $\stackrel{\sigma}{\rightarrow}$  are identical. With this convention, we can consider that the set of formulas is finite. It is denoted  $FORM^{\sigma}$ . Finally, in the following, *true* denotes any tautology and *false* denotes any inconsistant formula.

In the logical belief function theory, the user can assign masses to propositional formulas. This is shown by the following definition. **Definition 1** A logical mass function is a function  $m : FORM^{\sigma} \to [0, 1]$  such that:

$$m(false) = 0$$

and

$$\sum_{\varphi \in FORM^{\sigma}} m(\varphi) = 1$$

Like the belief function theory, the logical belief function theory offers several interesting functions for decision. In particular, the logical belief function and the logical plausibility functions are defined by:

**Definition 2** Given a logical mass function m, the logical belief function which is associated to m is defined by: Bel : FORM<sup> $\sigma$ </sup>  $\rightarrow$  [0,1] such that:

$$Bel(\varphi) = \sum_{\substack{\psi \in FORM^{\sigma} \\ \sigma \models \psi \to \varphi}} m(\psi)$$

**Definition 3** Given a logical mass function, the logical plausibility function which is associated to m is defined by:  $Pl: FORM^{\sigma} \rightarrow [0,1]$  such that

$$Pl(\varphi) = 1 - Bel(\neg \varphi)$$

It can be noticed that

$$Pl(\varphi) = \sum_{\substack{\psi \in FORM\sigma\\ (\sigma \land \psi \land \varphi) \text{ is satisfiable}}} m(\psi)$$

Like the belief function theory, the logical belief function theory offers several rules for combining logical mass functions. Let us only detail the logical DS rule defined by:

**Definition 4** Let  $m_1$  and  $m_2$  be two logical mass functions. The logical DS rule defines the logical mass function  $m_1 \oplus m_2 : FORM^{\sigma} \to [0, 1]$  by:

$$m_1 \oplus m_2(false) = 0$$

and for any  $\varphi$  such that  $\varphi \not \to false$ :

$$m_1 \oplus m_2(\varphi) = \frac{\sum_{(\varphi_1 \land \varphi_2) \stackrel{\sigma}{\leftrightarrow} \varphi} m_1(\varphi_1).m_2(\varphi_2)}{\sum_{\sigma \land \varphi_1 \land \varphi_2} is \ satisfiable} m_1(\varphi_1).m_2(\varphi_2)$$

if 
$$\sum_{\sigma \land \varphi_1 \land \varphi_2 \text{ is satisfiable}} m_1(\varphi_1).m_2(\varphi_2) \neq 0$$

It can easily be shown that this combination rule is the reformulation, into the logical belief function theory, of the Demspter's rule of combination.

#### 2.2 A trust model

The trust model we define here will be used to express the degrees at which an agent *i* thinks that another agent *j* is valid (resp is a misinformer, or complete, or is a falsifier) when it reports information  $\varphi$ . The model does not specify how these degrees are defined. For instance, they can be given by agent *i* on the basis on its past experience with *j*. More precisely, if *i* has already get information from *j*, it can evaluate how many times it believes it and how many times it believes the opposite. We can also imagine that if *i* does not know *j*, these degrees are provided by a reputation fusion mechanism or some other trust assessment model [15].

We consider a propositional language  $\Theta$  with two kinds of letters. The "information letters"  $p, q, \ldots$  will be used to model the information which is reported by the agents; the "reporting letters" of the form  $I_j\varphi$  will be used to represent "agent j reports information  $\varphi$ ", for any propositional formula  $\varphi$  made of information letters. For instance  $I_a p$  is a letter which represents "agent a reports information p",  $I_b(p \wedge q)$  is a letter which represents "agent b reports information  $p \wedge q$ ". **Definition 5** Consider two agents i and j and a piece of information  $\varphi$  i.e., a formula made of information letters. Let  $v_j \in [0,1]$  and  $m_j \in [0,1]$  two real numbers <sup>3</sup> such that  $0 \le v_j + m_j \le 1$ .  $v_j$  is the degree to which itrusts j for being valid relatively to  $\varphi$  and  $m_j$  is the degree to which i trusts j for being a misinformer relatively to  $\varphi$  (written  $VM(i, j, \varphi, v_j, m_j)$ ) iff i's beliefs can be modelled by the mass assignment  $m^{VM(i, j, \varphi, v_j, m_j)}$  defined by :

$$\begin{split} m^{VM(i,j,\varphi,v_j,m_j)}(I_j\varphi\to\varphi) &= v_j \\ m^{VM(i,j,\varphi,v_j,m_j)}(I_j\varphi\to\neg\varphi) &= m_j \\ m^{VM(i,j,\varphi,v_j,m_j)}(true) &= 1 - (v_j + m_j) \end{split}$$

According to this definition, if *i* believes at degree  $v_j$  that *j* is valid for  $\varphi$  and believes at degree  $m_j$  that *j* is a misinformer then its belief degree in the fact "if *j* reports  $\varphi$  then  $\varphi$  is true" is  $v_j$ ; its belief degree in the fact "if *j* reports  $\varphi$  then  $\varphi$  is true" is  $v_j$ ; its belief degree in the fact "if *j* reports  $\varphi$  then  $\varphi$  is false" is  $m_j$ ; and its total ignorance degree is  $1 - (v_j + m_j)$ . The following particular cases are worth pointing out:

- $(v_j = 1)$  and  $(m_j = 0)$  In this case,  $m^{VM(i,j,\varphi,1,0)}(I_j\varphi \to \varphi) = 1$ . I.e, *i* is certain that if *j* tells  $\varphi$  then  $\varphi$  is true, i.e, *i* is certain that *j* is valid for  $\varphi$ . I.e *i* totally trusts *j* for being valid relatively to  $\varphi$ .
- $(v_j = 0)$  and  $(m_j = 1)$  In this case,  $m^{VM(i,j,\varphi,0,1)}(I_j\varphi \to \neg\varphi) = 1$ . I.e. *i* is certain that if *j* reports  $\varphi$  then  $\varphi$  is false, i.e., *i* is certain that *j* is a *misinformer* for  $\varphi$ . I.e *i* totally trusts *j* for being misinformer relatively to  $\varphi$ .

**Definition 6** Consider two agents i and j and a piece of information  $\varphi$ . Let  $c_j \in [0,1]$  and  $f_j \in [0,1]$  two real numbers<sup>4</sup> such that  $0 \le c_j + f_j \le 1$ .  $c_j$  is the degree to which i trust j for being complete relatively to  $\varphi$  and  $f_j$  is the degree to which i trusts j for being a falsifier relatively to  $\varphi$  (written  $CF(i, j, \varphi, c_j, f_j)$ ) iff i's beliefs can be modelled by the mass assignment  $m^{CF(i,j,\varphi,c_j,f_j)}$  defined by:

$$\begin{split} m^{CF(i,j,\varphi,c_j,f_j)}(\varphi \to I_j\varphi) &= c_j \\ m^{CF(i,j,\varphi,c_j,f_j)}(\neg \varphi \to I_j\varphi) &= f_j \\ m^{CF(i,j,\varphi,c_j,f_j)}(true) &= 1 - (c_j + f_j) \end{split}$$

According to this definition, if *i* believes at degree  $c_j$  that *j* is complete for  $\varphi$  and believes at degree  $f_j$  that *j* is a falsifier then its belief degree in the fact "if  $\varphi$  is true then *j* reports  $\varphi$ " is  $c_j$ ; its belief degree in the fact "if  $\varphi$  is false then *j* reports  $\varphi$ " is  $f_j$ ; and its total ignorance degree is  $1 - (c_j + f_j)$ . The following particular cases are worth pointing out:

- $(c_j = 1)$  and  $(f_j = 0)$  In this case,  $m^{CF(i,j,\varphi,1,0)}(\varphi \to I_j\varphi) = 1$ . I.e, *i* is certain that if  $\varphi$  is true then *j* reports  $\varphi$  i.e, *i* is certain that *j* is complete for  $\varphi$ . I.e *i* totally trusts *j* for being complete relatively to  $\varphi$ .
- $(c_j = 0)$  and  $(f_j = 1)$  In this case,  $m^{CF(i,j,\varphi,0,1)}(\neg \varphi \rightarrow I_j \varphi) = 1$ . I.e. *i* is certain that *j* is a falsifier for  $\varphi$ . I.e *i* totally trusts *j* for being falsifier relatively to  $\varphi$ .

## 2.3 A quick comparision with graded trust model

According to Demolombe, [8], the notion of graded trust envolves two components respectively called graded beliefs and graded regularities. Consequently, graded trust are modelled by formulas of the form:  $B_i^g(\phi \to^h \psi)$  where  $B_i^g$  is a modal operator so that  $B_i^g a$  expresses that the strength of agent *i*'s beliefs in *a* is *g* and  $\to^h$  is a conditional so that  $(\phi \to^h \psi)$  expresses that  $\phi$  entails  $\psi$  at degree *h*.

The trust model we defined in the previous section does not permit to quantify the strength of the implication. So Demolombe's graded trust model is richer. For instance it can represent the fact that agent i is totally certain that j is highly valid for proposition p. The previous model cannot.

As for graded beliefs, the two models provide different ways to model them since they do not impose the same axioms on graded beliefs. In particular, in the logical belief function theory, we have: if  $Bel_i(\varphi_1) = g_1$  and  $Bel_i(\varphi_2) = g_2$  then  $Bel_i(\varphi_1 \land \varphi_2) \leq min(g_1, g_2)$  and  $Bel_i(\varphi_1 \lor \varphi_2) \geq max(g_1, g_2)$ . These two assertions are less restrictive that axioms (U2) and (U3) of [8] which impose that  $Bel_i(\varphi_1 \land \varphi_2) = min(g_1, g_2)$  and  $Bel_i(\varphi_1 \lor \varphi_2) = max(g_1, g_2)$ .

<sup>&</sup>lt;sup>3</sup>These degrees should be indexed by i and by  $\varphi$  but this is omitted for readibility

<sup>&</sup>lt;sup>4</sup>Again these degrees should be indexed by i and  $\varphi$ 

# 3 Applying this trust model to evaluate information

The question which is addressed in this section is:

To which extent an agent can believe information he gets?

We will successively examine the case when the agent directly gets information from the information source, then the case when there is an intermediary agent between the agent and the information source.

#### 3.1 The agent is in direct contact with the source

Let us first give the following preliminary definitions.

**Definition 7**  $m^{VMCF}$  denotes the mass assignment obtained by combining the two previous mass assignments. *I.e.*,

$$m^{VMCF} = m^{VM(i,j,\varphi,v_j,m_j)} \oplus m^{CF(i,j,\varphi,c_j,f_j)}.$$

This assignment represents i's degrees of trust in the fact that j is valid, complete, misinformer or falsifier relatively to information  $\varphi$ . One can check that  $m^{VMCF}$  is defined by:

$$\begin{split} m^{VMCF}(\varphi\leftrightarrow I_{j}\varphi) &= v_{j}.c_{j} \\ m^{VMCF}(\varphi) &= v_{j}.f_{j} \\ m^{VMCF}(I_{j}\varphi\rightarrow\varphi) &= v_{j}.(1-c_{j}-f_{j}) \\ m^{VMCF}(\neg\varphi) &= m_{j}.c_{j} \\ m^{VMCF}(\neg\varphi\leftrightarrow I_{j}\varphi) &= m_{j}.f_{j} \\ m^{VMCF}(I_{j}\varphi\rightarrow\neg\varphi) &= m_{j}.(1-c_{j}-f_{j}) \\ m^{VMCF}(\varphi\rightarrow I_{j}\varphi) &= (1-v_{j}-m_{j}).c_{j} \\ m^{VMCF}(\neg\varphi\rightarrow I_{j}\varphi) &= (1-v_{j}-m_{j}).f_{j} \\ m^{VMCF}(true) &= (1-v_{j}-m_{j}).(1-c_{j}-f_{j}) \end{split}$$

**Definition 8**  $m^{\psi}$  is the mass assignments defined by:  $m^{\psi}(\psi) = 1$ .

In particular, if  $\psi$  is  $I_j \varphi$ , then  $m^{I_j \varphi}$  represents the fact that agent *i* is certain that *j* has reported  $\varphi$ . If  $\psi$  is  $\neg I_j \varphi$ , then  $m^{\neg I_j \varphi}$  represents the fact that agent *i* is certain that *j* did not report  $\varphi$ .

#### 3.1.1 First case.

Here, we assume that agent i get information from the information source j which reports  $\varphi$ . In this case, i's beliefs are modelled by the following mass assignment  $m_i$ :

$$m_i = m^{VMCF} \oplus m^{I_j\varphi}$$

One can check that  $m_i$  is defined by:

$$m_i(I_j\varphi \wedge \varphi) = v_j$$
  

$$m_i(I_j\varphi \wedge \neg \varphi) = m_j$$
  

$$m_i(I_j) = (1 - v_j - m_j)$$

**Theorem 1** Let  $Bel_i$  be the belief function associated with assignment  $m_i$ . Then:

$$Bel_i(\varphi) = v_i, Bel_i(\neg \varphi) = m_i$$

Consequently, when *i* knows that *j* reported  $\varphi$  and when  $VM(i, j, \varphi, v_j, m_j)$  and  $CF(i, j, \varphi, c_j, f_j)$ , then *i* believes  $\varphi$  more than  $\neg \varphi$  if and only if  $v_j > m_j$  i.e, its belief degree in *j*'s being valid is greater that its belief degree in *j*'s being a misinformer. These results are not surprising.

#### Example 1

In this example, we consider an agent (denoted a) which gets information provided by Météo-France web site (denoted MF).  $\Theta$  is the propositional language whose "information letters" are: rain (tomorrow will be rainy), cold (tomorrow will be cold), dry (tomorrow will be dry), warm (tomorrow will be warm), and whose "reporting letters" are:  $I_{MF}$  rain (Météo-France forecats rain for tomorrow),  $I_{MF}$  rain (Météo-France forecats no rain for tomorrow), etc.

 $\sigma$  is here the formula  $\sigma = \neg(rain \land dry) \land \neg(cold \land warm)$ 

1. Suppose that Météo-France web site indicates that tomorrow will be rainy. Suppose that, given a's numerous previous access to this web site, a believes at degree 0.8 that Météo-France web site is valid for rain forecast and a believes at degree 0.1 that it misinforms for rain forecast i.e., VM(a, MF, rain, 0.8, 0.1).

Then, by theorem 1, we have:  $Bel_a(rain) = 0.8$  and  $Bel_a(\neg rain) = 0.1$ , i.e., a believes at degree 0.8 that tomorrow will be rainy and a believes at degree 0.1 that tomorrow will be dry. Notice that we also have  $Bel_a(\neg dry) = 0.8$  and  $Bel_a(dry) = 0.1$ .

- 2. Now suppose that Météo-France web site forecats rain and cold for tomorrow. Suppose that a trusts this web site at degree 0.8 for being valid and a trusts it at degree 0.1 for being misinformer, i.e  $VM(a, MF, (rain \land cold), 0.8, 0.1)$ . Then, by theorem 1, we have:
  - $Bel_a(rain \wedge cold) = Bel_a(rain) = Bel_a(cold) = 0.8.$
  - $Bel_a(\neg rain \lor \neg cold) = Bel_a(dry \lor warm) = 0.1,$
  - $Bel_a(\neg rain) = Bel_a(\neg cold) = Bel_a(dry) = Bel_a(warm) = 0.$

*I.e.*, *a's* degree of belief in the fact that tomorrow will be rainy and cold (resp tomorrow will be rainy, tomorrow will be cold) is 0.8. *a's* degree of belief in the fact that tomorrow will be dry or warm is 0.1. But a's degree of belief in that the fact that tomorrow will be dry (resp will be warm) is 0.

One can notice however that we also have  $Pl_a(dry) = Pl_a(warm) = 0.2$  i.e., a's plausibility degree in that the fact that tomorrow will dry (resp will be warm) is 0.2.

As a consequence of theorem 1 we have:

- If  $VM(i, j, \varphi, 1, 0)$  then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$  i.e., *i* believes  $\varphi$  and does not believe  $\neg \varphi$ ;
- If  $VM(i, j, \varphi, 0, 1)$  then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$  i.e., *i* does not believe  $\neg \varphi$  and believes  $\varphi$ .

The first result is coherent with a result obtained in Demolombe's model in which  $T_{valid_{i,j}}(\varphi) \wedge B_i I_J^i \varphi \to B_i \varphi$ and  $T_{valid_{i,j}}(\varphi) \wedge B_i I_J^i \varphi \to \neg B_i \neg \varphi$  are theorems. However, Demolombe's model does not model misinformers thus the second result cannot be compared.

### 3.1.2 Second case.

Here we assume that the information source j did not report  $\varphi$ . In this case, i's beliefs can be modelled by the mass assignment  $m_i$ :

$$m_i = m^{VMCF} \oplus m^{\neg I_j \varphi}$$

One can check that  $m_i$  is defined by:

$$\begin{split} m_i(\neg I_j\varphi \wedge \varphi) &= f_j \\ m_i(\neg I_j\varphi \wedge \neg \varphi) &= c_j \\ m_i(\neg I_j) &= (1-c_j-f_j) \end{split}$$

**Theorem 2** Let  $Bel_i$  be the belief function associated with assignment  $m_i$ . Then,

$$Bel_i(\varphi) = f_j, Bel_i(\neg \varphi) = c_j$$

Consequently, when i knows that j did report  $\varphi$  and when  $VM(i, j, \varphi, v_j, m_j)$  and  $CF(i, j, \varphi, c_j, f_j)$  then i believes  $\varphi$  more than  $\neg \varphi$  if and only if  $f_j > c_j$  i.e., its belief degree in j's being a falsifier is greater that its belief degree in j's being complete. Again this is not surprising.

**Example 2** Suppose now that agent a reads Météo-France web site to be informed about storms in the south of France and that no forecast of storm is indicated on the site. Here we consider a propositional language whose "information letter" are: storm (there will be a storm in the south of France), beach (beach access will be open) and whose "reporting letters" are:  $I_{MF}$  storm (Météo-France forecasts a storm in the south of France),  $I_{MF}$ -storm (Météo-France forecasts no storm in the south of France).

Consider that  $\sigma = storm \rightarrow \neg beach$ , i.e, in case of storm beach access will be closed.

Suppose that a trusts at degree 0.8 this web site for being complete for storm forecast and a trusts at degree 0.1 this site for being falsifier for storm forecast, i.e., CF(a, MF, storm, 0.8, 0.1).

Then, according to theorem 2, we have:  $Bel_a(storm) = 0.1$  and  $Bel_a(\neg storm) = 0.8$ . Consequently, a has a greater degree of belief in the fact that there will be no storm than in the fact that there will be one. We also have  $Bel_a(\neg beach) = 0.1$  and  $Bel_a(beach) = 0$ .  $Bel_a(beach) = 0$  may look strange but it comes from the fact that a has no guarantee that beach acces will be open (even if there is no storm). Indeed,  $\neg$ storm  $\rightarrow$  beach is not deducible from  $\sigma$ , i.e., nothing states that if there is no storm, beach access will be open.

As a consequence of theorem 2 we have:

- If  $CF(i, j, \varphi, 1, 0)$ , then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$  i.e., i does not believes  $\varphi$  and believes  $\neg \varphi$ ;
- If  $CF(i, j, \varphi, 0, 1)$ , then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$  i.e., *i* believes  $\varphi$  and does not believe  $\neg \varphi$ .

The first result is coherent with a result obtained in Demolombe's model in which  $T_{complete_{i,j}}(\varphi) \wedge B_i \neg I_J^i \varphi \rightarrow B_i \neg \varphi$  and  $T_{complete_{i,j}}(\varphi) \wedge B_i \neg I_J^i \varphi \rightarrow \neg B_i \varphi$  are theorems. However, Demolombe's model does not model falsifiers thus the second result cannot be compared.

#### 3.2 There is a third agent between the agent and the source

We consider here that i is not in direct contact with the source of information k, but there is a go-between agent named j. The question is again to know to which extent agent i can believe information it gets. Here, we consider a propositional language whose "information letters" are: p, q... and whose "reporting letters" are  $I_k\varphi, I_jI_k\varphi, I_j\neg I_k\varphi$ . which respectively mean k reported  $\varphi, j$  told that k reported  $\varphi, j$  told that k did not report  $\varphi$ .

In this section, the mass assignment  $m^{VMCF}$  is defined by:

#### **Definition 9**

$$m^{VMCF} = m^{VM(i,j,I_k\varphi,v_j,m_j)} \oplus m^{CF(i,j,I_k\varphi,c_j,f_j)} \oplus m^{VM(i,k,\varphi,v_k,m_k)} \oplus m^{CF(i,k,\varphi,c_k,f_k)}$$

This assignment represents the beliefs of i as regard to agent j being valid, complete, a misinformer or a falsifier for information  $I_k\varphi$  and as regard to agent k being valid, complete, a misinformer or a falsifier for information  $\varphi$ .

#### 3.2.1 First case.

Assume that j reported that k told it  $\varphi$ . In this case, i's beliefs are modelled by the mass assignment  $m_i$  defined by:

$$m_i = m^{VMCF} \oplus m^{I_j I_k \varphi}$$

**Theorem 3** Let  $Bel_i$  be the belief function associated with assignment m.

$$Bel_i(\varphi) = v_k \cdot f_k + v_k \cdot v_j - v_k \cdot f_k \cdot v_j - v_k \cdot f_k \cdot m_j + f_k \cdot m_j \quad and \\Bel_i(\neg \varphi) = m_k \cdot c_k + m_k \cdot v_j - m_k \cdot c_k \cdot v_j + c_k \cdot m_j - m_k \cdot c_k \cdot m_j$$

Consequently,

- If  $VM(i, j, I_k\varphi, 1, 0)$  and  $VM(i, k, \varphi, 1, 0)$  then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$ .
- If  $VM(i, j, I_k\varphi, 1, 0)$  and  $VM(i, k, \varphi, 0, 1)$  then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$ .
- If  $VM(i, j, I_k\varphi, 0, 1)$  and  $CF(i, k, \varphi, 1, 0)$  then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$ .
- If  $VM(i, j, I_k\varphi, 0, 1)$  and  $CF(i, k, \varphi, 0, 1)$  then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$ .

The first result could be provided in Demolombe's model if the inform operator was defined by omitting the agent which receives the information. In this case, we would have  $Tvalid_{i,k}(\varphi) = B_i(I_k\varphi \to \varphi)$ . Thus  $Tvalid_{i,j}(I_k\varphi) \wedge Tvalid_{i,k}(\varphi) \wedge I_jI_k\varphi \to B_i\varphi$  and  $Tvalid_{i,j}(I_k\varphi) \wedge Tvalid_{i,k}(\varphi) \wedge I_jI_k\varphi \to \neg B_i\neg\varphi$  would be theorems. As for the last three results, they cannot be compared since Demolombe's framework does not model misinformers nor falsifiers.

## 3.2.2 Second case.

Here, we assume that j does not tell that k reported  $\varphi$ . In this case, i's beliefs are modelled by:

$$m_i = m^{VMCF} \oplus m^{\neg I_j I_k \varphi}$$

**Theorem 4** Let  $Bel_i$  be the belief function associated with assignment m.

$$Bel_i(\varphi) = v_k.c_k.f_j + v_k.f_k + v_k.(1 - c_k - f_k).(1 - c_j) + m_k.f_k.c_j + (1 - v_k - m_k).f_k and Bel_i(\neg \varphi) = v_k.c_k.c_j + m_k.c_k + m_k.f_k.(1 - c_j) + m_k.(1 - c_k - f_k).f_j + (1 - v_k - m_k).c_k.c_j$$

Consequently,

- If  $CF(i, j, I_k\varphi, 1, 0)$  and  $CF(i, k, \varphi, 1, 0)$  then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$ .
- If  $CF(i, j, I_k\varphi, 1, 0)$  and  $CF(i, k, \varphi, 0, 1)$  then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$ .
- If  $CF(i, j, I_k\varphi, 0, 1)$  and  $VM(i, k, \varphi, 1, 0)$  then  $Bel_i(\varphi) = 1$  and  $Bel_i(\neg \varphi) = 0$ .
- If  $CF(i, j, I_k\varphi, 0, 1)$  and  $VM(i, k, \varphi, 0, 1)$  then  $Bel_i(\varphi) = 0$  and  $Bel_i(\neg \varphi) = 1$ .

Again, the first result could be provided in Demolombe's model if the inform operator was defined by omitting the agent which receives the information. In this case, we would have  $Tcomplete_{i,k}(\varphi) = B_i(\varphi \to I_j\varphi)$ . Thus  $Tcomplete_{i,j}(I_k\varphi) \wedge Tcomplete_{i,k}(\varphi) \wedge \neg I_j I_k\varphi \to B_i \neg \varphi$  and  $Tvalid_{i,j}(I_k\varphi) \wedge Tvalid_{i,k}(\varphi) \wedge I_j I_k\varphi \to \neg B_i\varphi$  would be theorems. As for the last three results, they cannot be compared since Demolombe's framework does not model misinformers nor falsifiers.

The following example illustrates the first item.

**Example 3** Let us consider three agents denoted me (me), n (my neighbour) and MF (Météo-France web site). Suppose that my neighbour, who regularly reads Météo-France web site, does not tell me that a storm is forecasted in the south of France.

Suppose that I trust my neighbour to be complete (he always tell me what he reads on Météo-France web site) i.e. in particular we have  $CF(me, n, I_{MF}storm, 1, 0)$ .

Suppose that I also trust Météo-France to be complete relatively to storms forecats (they always indicate forecasted storms) i.e., we have: and CF(me, MF, storm, 1, 0).

In this case, we get  $Bel_{me}(storm) = 0$  and  $Bel_{me}(\neg storm) = 1$  i.e., I can conclude that there will be no storm.

#### 3.2.3 Third case

Here, we assume that j reports that k did not report  $\varphi$ . In this case, i's beliefs are modelled by:

$$m_i = m^{VMCF} \oplus m^{I_j \neg I_k \varphi}$$

We do not detail this case but we think that the reader gets the idea.

#### 3.2.4 Fourth case

Here, we assume that j does not tell that k did not report  $\varphi$ . In this case, i's beliefs are modelled by:

$$m_i = m^{VMCF} \oplus m^{\neg I_j \neg I_k \varphi}$$

We do not detail this case but the reader can easily imagine the kind of conclusions we can derive.

## 4 Concluding remarks

To which extent an agent can believe a piece of information it gets from an information source? This was the question addressed in this paper in which we have provided a model for expressing the relations between the trust an agent puts in the sources and its beliefs about the information they provide. This model is based on Demolombe's model and extends it by considering information sources that can report false information. Furthermore, this model is defined in the logical belief function theory allowing degrees of trust to be modelled in a quantitative way and allowing the agent to consider integrity constraints. We have shown that not only this model can be used when the agent directly gets information from the source but it can also be used when the agent gets second hand information i.e., when the agent is not directly in contact with the source.

Notice that the belief function theory has already been used in problems related to trust management. For instance, subjective logic, a specific case of belief function theory, has been used in [11] for trust network analysis. Notice however that trust network analysis is not the problem addressed here. [6], already mentionned in the introduction is another paper which uses belief function for estimating the plausibility of a piece of information emitted by sources which may be unrelevant or untruthful. But the case of second hand information is not addressed. More recently [17] defines a formalism based on description logic and belief function theory to reason about uncertain information provided by untrustworthy sources. However, in this work, a source is given a single degree. This degree, called the "degree of trust", is used for discounting the information it provides. It looks close to the "degree at which an agent trusts a source for being valid" as introduced here but a formal comparision has to be done. Furthermore again, the case of second hand information is not addressed.

As to the use of the logical belief function theory, a question concerns the choice of the combination rule. Here, we have chosen the logical DS rule of combination which is the reformulation of the most classical rule. But can the other rules of combination be used? Or more precisely, in which cases, should we use them instead? This is an important question to be answered in the future.

The model defined here would be more general if it could define information sources properties according to sets of propositions and not to a given proposition. For instance, we could express that an agent is valid for any proposition related to the topic "weather forecasts" and complete for any proposition related to the topic "south of France". That would imply that this agent is valid and complete for any information related to weather forecast in the south of France. However, such a reasoning has to be formally defined and describing these topics by means of an ontology is a possible solution.

Finally, this work could be extended by considering the case when information sources provide uncertain information like [17] does. This would allow us to deal with the case when an agent reports that some fact is highly certain or when an agent reports that another agent has reported that some fact was fairly certain. Modelling this kind of uncertainty and mixing it with the one which is introduced in this paper would allow us to deal with more complex cases.

Acknowledgements. I thank the anonymous referees whose pertinent questions led me to improve this paper.

## References

 C. E. Alchourron, P. Gardenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic, 50:510530, 1985.

- [2] B. F Chellas. Modal logic: An introduction. Cambridge University Press, 1980.
- [3] J. Chauvet-Paz and L. Cholvy. Reasoning with Information reported by Imperfect Agents. Proceedings of First International Workshop on Uncertainty Reasoning and Multi-Agent Systems for Sensor Networks (URMASSN'11), Belfast, Northern Ireland, 2011.
- [4] L. Cholvy. Collecting information reported by imperfect information sources. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, R. R. Yager (Eds.): Advances in Computational Intelligence: 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2012), Springer, 2012,
- [5] L. Cholvy. Logical representation of beliefs in the belief function theory. Proceedings of IJCAI 2013 Workshop on Weighted Logics for Artificial Intelligence. August 2-3 2013, Beijing, China.
- [6] F. Pichon, D. Dubois, T. Denoeux. Relevance and Truthfulness in Information Correction and Fusion. International Journal of Approximate Reasoning, 53(2), 159-175 (2012).
- [7] R. Demolombe, *Reasoning about trust: a formal logical work.* in: Proceedings of the 2<sup>nd</sup> second International Conference iTrust, 291-303, 2004.
- [8] R. Demolombe. Graded Trust. In Proceedings of Workshop on Trust in Agent Societies, Budapest, 2009.
- [9] R. Demolombe. Transitivity and Propagation of Trust in Information Sources. An Analysis in Modal Logic In Proceedings of the 12th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA 2011), LNAI 6814, Springer-Verlag, 2011.
- [10] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. Artificial Intelligence, 172 (23), 234264, 2008.
- [11] A Josang, R. Hayward and S. Pope. Trust Network Analysis with subjective logic In Proceedings of 29<sup>th</sup> Australasian Computer Science Conference (ACSC2006), 2006.
- [12] H. Katsuno and A. Mendelzon. On the Difference between Updating a Knowledge Base and Revising it. Proceedings of Principles of Knowledge Representation and Reasoning (KR), 1991.
- [13] H. Katsuno and A. O. Mendelzon Propositional knowledge base revision and minimal change. Artificial Intelligence, 52:263294. 1991
- [14] C-J. Liau Belief, information acquisition, and trust in multi-agent systems. A modal logic formulation. Artificial Intelligence, 149:31-60. 2003.
- [15] TD Huynh, N Jennings, N Shadbolt. FIRE: An integrated trust and reputation model for open multiagent systems. *Proceedings of 16th European Conference on Artificial Intelliegnce, ECAI 2004.*
- [16] E. Lindhal, S. O'Hara, Q. Zhu. A Multi-Agent System of Evidential Reasoning for Intelligence Analyses In Edmund H. Durfee, Makoto Yokoo, Michael N. Huhns, Onn Shehory (Eds.): Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007), Honolulu, Hawaii, USA, May 14-18, 2007
- [17] M. Sensoy et al. Reasoning aabout Uncertain Information and Conflict Resolution through Trust Revision In M. L. Gini, O. Shehory, T. Ito, C M. Jonker (Eds.). Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013), Saint Paul, MN, USA, May 6-10, 2013.
- [18] G. Shafer. A mathematical Theory of Evidence Princeton University Press, 1976.
- [19] STANAG 2511 Standardization Agreement 2511. Intelligence reports, NATO (unclassified), january 2003.
- [20] M. Winslett Updating Logical Databases. Cambridge University Press, 1990.