

Chapter 1

TO TRUST INFORMATION SOURCES: A PROPOSAL FOR A MODAL LOGICAL FRAMEWORK

Robert Demolombe
ONERA
Centre de Toulouse. DTIM
France
Robert.Demolombe@cert.fr

1. INTRODUCTION

Different meanings can be attributed to the concept of trust (see Lerch and Prietula, 1989; P.Oerbaek, 1995; March, 1994; Thompson, 1984). We can understand trust as an attitude of an agent who believes that another agent has a given property. So, we can analyse the meaning of trust as function of the attributed property.

For instance, the property may be that the agent we trust fulfils his obligations, as it is the case for a notary, or any kind of trusted third party. It may also be that he has the ability to perform a given kind of action, as it is the case for a bodyguard. Here, the property we consider is that the agent is an information source who delivers correct information, as it may be the case, for instance, in the context of an intelligence service.

In this paper we start by analysing what it means that an agent “attributes” a property to another agent, and that an agent “delivers correct information”. This analysis is based on four elementary properties that are called “sincerity”, “credibility”, “cooperativity” and “vigilance”. In section 2 these properties are informally defined. In section 3, we propose formal definitions in modal logic (see Chellas, 1988) of trust in reference to one of these properties, and we introduce a logic for reasoning about consequences of a set of assumptions which represent a given situation. Examples of reasoning are given in section 4. Finally, an extension to graded trust is presented in section 5.

2. BASIC CONCEPTS

We consider contexts where several information sources, called “agent”, can communicate with one another, and where it is known that some agents may not be reliable. Then, it is important to be able to characterise information sources who can be trusted.

As we have said before, we do not trust an agent for everything and we have to define the property which is implicitly referred to. We also have to define what it means that this property is “attributed” to an agent.

In our work we consider that this means that the agent a who attributes this property to another agent b “strongly believes” that this property is held by b ¹. We do not say that agent a knows that the property is held by b because we are in a context where no information can be taken as true, in an absolute sense. We assume that when agent a strongly believes some proposition represented by p , a strongly believes that, if he strongly believes that p is true then p is true.

A consequence of this definition of strong belief is that if a strongly believes p , and another agent told him that p is false, agent a will reject the fact that it is false, and he will not change his mind. According to this definition, if an agent a trusts b in regard to something, then a has an irrevocable trust in b for that thing.

Now, we can introduce the different properties involved in different definitions of trust.

Sincerity

An agent a is sincere with respect to agent b when he believes what he says to b . In other words, if a has informed another agent b about p , then a believes p . Here, the fact that a has informed b means that a has performed an action whose effect is that b knows that a wanted to inform him about p .

It is worth noting that the sincerity property has a conditional form. That is because this property does not depend on a particular situation. It may be that in some situation a says nothing about p , and, in another situation, he says that p is true. The fact that he is sincere means that in **any** situation if he says p he believes p . Notice also that this definition of sincerity is quite restrictive since it is restricted to a given proposition. So, it may be assumed that a is sincere for p and not sincere for q .

Credibility

We say that an agent a is credible when what he believes is true in the world. In fact this definition is also restricted to a given proposition. So, a is credible for p is taken in the sense that if a believes p then p is

true. The credibility definition has also a conditional form, and it holds in a set of situations.

Cooperativity

Cooperativity is defined by duality of sincerity. An agent a who is assumed to be cooperative with respect to agent b is an agent who says what he believes. To be more precise, in our definition, the fact that agent a is cooperative with respect to agent b for p means that if agent a believes p then a informs b about p . Intuitively, a does not want to keep for himself the fact that he believes p .

Vigilance

Vigilance is also defined by duality of credibility. In our definition, a is vigilant for p means that if p is true in the world then a believes p . In other terms, it cannot be the case that p is true and a does not believe p , that is, if p was false, at the moment p comes to be true, a believes p .

Validity

We define validity property as the conjunction of sincerity and credibility. That is, agent a has property of validity for p with respect to agent b if and only if a is sincere with respect to agent b for p and a is credible for p . A consequence of this definition is that if a is valid for p , if a informs b about p , then p is true.

Completeness

We define completeness property as the conjunction of cooperativity and vigilance. That is, agent a has property of completeness for p with respect to agent b if and only if a is cooperative with respect to agent b for p and a is vigilant for p . A consequence of this definition is that if a is complete for p , if p is true, then a informs b about p .

There are application domains where information sources may be artificial agents, like sensors or database systems. In that cases, it is not clear that properties of sincerity and cooperativity have an intuitive meaning. Then, validity and completeness can be taken as primitive concepts.

Now, we can define trust by combining the notion of strong belief and one of the properties presented before. For instance, we say that a trusts b for the sincerity of b with respect to a for p if and only if a strongly believes that b is sincere with respect to a for p . In a similar way, we say that a trusts b for credibility of b for p if and only if a strongly believes that b is credible for p .

3. FORMAL DEFINITION OF CONCEPTS

To give formal definitions of basic concepts we use three modal operators. They are intended to formalise the concepts of belief, strong belief

and information action. Sentences in the scope of modal operators can be sentences of propositional calculus with nested modal operators.

Belief

The fact that agent a believes a proposition represented by p is denoted by $B_a p$. It is assumed that a set of agent's beliefs is consistent, and that an agent has reasoning capacities compatible with modus ponens. We have the axiom schemas:

$$(K1) \quad B_a(p \rightarrow q) \rightarrow (B_a p \rightarrow B_a q)$$

$$(D1) \quad \neg(B_a p \wedge B_a \neg p)$$

Strong belief

The fact that agent a strongly believes a proposition represented by p is denoted by $K_a p$. We also assume that the K_a operator obeys (K) and (D) axiom schemas. In addition we assume that it obeys the (KT) axiom schema.

$$(K2) \quad K_a(p \rightarrow q) \rightarrow (K_a p \rightarrow K_a q)$$

$$(D2) \quad \neg(K_a p \wedge K_a \neg p)$$

$$(KT) \quad K_a(K_a p \rightarrow p)$$

The schema (KT) explicitly states that property $K_a p \rightarrow p$ is not guaranteed to be true, but it is in a 's strong belief. To express the fact that strong beliefs are a special kind of beliefs, we have the following axiom schema:

$$(KB) \quad K_a p \rightarrow B_a p$$

For operators B_a and K_a we accept the inference rule of necessitation.

Information action

The fact that agent a has informed agent b about p is denoted by $I_{a,b}(p)$, or by $I_{a,b}p$, for simplicity. It is assumed that the choice of the particular sentence p used to represent a proposition is irrelevant. That is, sentence p can be substituted by any other sentence q logically equivalent to p . Then, we have the inference rule:

$$(RE) \quad \frac{\vdash p \leftrightarrow q}{\vdash I_{a,b}p \leftrightarrow I_{a,b}q}$$

Presently, (RE) is the only property we have accepted for the $I_{a,b}p$ operator. That means that we do not consider that the two situations

represented by $I_{a,b}(p \wedge q)$ and by $I_{a,b}p$ and $I_{a,b}q$ have necessarily to be considered as equivalent. This is an open question which requires further investigations.

We also have assumed that when two agents exchange information there is no failure in the communication process. That is formally represented by the two axiom schemas (OBS1) and (OBS2) whose intuitive meaning is that agent a does not ignore whether an other agent b has informed him about some sentence p . So, we have:

$$(OBS1) \quad I_{b,a}p \rightarrow K_a(I_{b,a}p)$$

$$(OBS2) \quad \neg I_{b,a}p \rightarrow K_a(\neg I_{b,a}p)$$

In these axiom schemas we use operator K_a because agent a observes the situation and he has no doubt about actions which have, or have not, been performed.

We can now formally define the different sorts of trust. The fact that a trusts b for p in regard to sincerity (respectively credibility, cooperativity, vigilance, validity, completeness) is denoted by $Tsinc_{a,b}(p)$ (respectively $Tcred_{a,b}(p)$, $Tcoop_{a,b}(p)$, $Tvigi_{a,b}(p)$, $Tval_{a,b}(p)$, $Tcomp_{a,b}(p)$) and is formally defined by:

$$Tsinc_{a,b}(p) \stackrel{\text{def}}{=} K_a(I_{b,a}p \rightarrow B_b p)$$

$$Tcred_{a,b}(p) \stackrel{\text{def}}{=} K_a(B_b p \rightarrow p)$$

$$Tcoop_{a,b}(p) \stackrel{\text{def}}{=} K_a(B_b p \rightarrow I_{b,a}p)$$

$$Tvigi_{a,b}(p) \stackrel{\text{def}}{=} K_a(p \rightarrow B_b p)$$

$$Tval_{a,b}(p) \stackrel{\text{def}}{=} Tsinc_{a,b}(p) \wedge Tcred_{a,b}(p)$$

$$Tcomp_{a,b}(p) \stackrel{\text{def}}{=} Tvigi_{a,b}(p) \wedge Tcoop_{a,b}(p)$$

Since K_a is a normal modal operator we have $Tval_{a,b}(p)$ (respectively $Tcomp_{a,b}(p)$) implies $K_a(I_{b,a}p \rightarrow p)$ (respectively $K_a(p \rightarrow I_{b,a}p)$).

4. REASONING ABOUT TRUST

The logic and formal definitions of concepts presented in the previous section allow us to represent a given situation, and to derive consequences from this representation.

A situation may be formally defined as a set S of sentences of the form $Tprop_{a,b}(p)$, where $prop$ may be one of the above properties, and a set of sentences of the form $I_{b,a}p$ or $\neg I_{b,a}p$. These kinds of sentences can be used, in particular, to represent trusted information sources and communication action which have been performed.

We denote by \vdash_T the consequence relation for the logic we have defined. To analyse properties of a given situation S we have to prove that a given sentence $cons$ is a consequence of S , that is:

$$\vdash_T S \rightarrow cons$$

In this context we are mainly interested by consequences that represent agent's beliefs or agent's strong beliefs, that is, consequences of the form $K_a p$ or $B_a p$. Let us illustrate this with a toy example.

We consider three agents a , b and c who are interested to exchange information about the two facts "there is a spy in the train T31", denoted by p , and "the train T31 has arrived at the railway station", denoted by q . In this situation agent a trusts b in regard to his validity for p , and in regard to his sincerity for q , and a trusts c in regard to his completeness for q . a 's trust may be supported, for instance, by the fact that b belongs to some intelligence service, and c is an employee of the railway station who stands on the platform where the train is supposed to arrive. This is formally represented by $Tval_{a,b}(p)$, $Tsinc_{a,b}(q)$ and $Tcomp_{a,c}(q)$. In that situation b has transmitted to a information p , and he has also transmitted q , and c has not transmitted to a information q . This is formally represented by $I_{b,a}p$, $I_{b,a}q$ and $\neg I_{c,a}q$. So, we have:

$$S = \{Tval_{a,b}(p), Tsinc_{a,b}(q), Tcomp_{a,c}(q), I_{b,a}p, I_{b,a}q, \neg I_{c,a}q\}$$

We want to know what are a 's beliefs and strong beliefs about the fact that there is a spy in the train T31, and about the fact that the train has arrived in the railway station. In our logic we can draw the following consequences from S .

- | | | |
|-----|-------------------------------|------------------------------|
| (1) | $I_{b,a}p$ | in S |
| (2) | $K_a(I_{b,a}p)$ | (1) and (OBS1) |
| (3) | $Tval_{a,b}(p)$ | in S |
| (4) | $K_a(I_{b,a}p \rightarrow p)$ | (3) and definition of $Tval$ |

(5)	$K_a p$	(2), (4) and (K2)
(6)	$\neg I_{c,a} q$	in S
(7)	$K_a (\neg I_{c,a} q)$	(6) and (OBS2)
(8)	$Tcomp_{a,c}(q)$	in S
(9)	$K_a (q \rightarrow I_{c,a} q)$	(8) and definition of $Tcomp$
(10)	$(q \rightarrow I_{c,a} q) \rightarrow$ $(\neg I_{c,a} q \rightarrow \neg q)$	tautology
(11)	$K_a (\neg I_{c,a} q \rightarrow \neg q)$	(9), (10) and properties of K_a
(12)	$K_a (\neg q)$	(7), (11) and (K2)
(13)	$K_a (p \wedge \neg q)$	(5), (12) and properties of K_a

It is interesting to see in this derivation how sentence (12) $K_a(\neg q)$ is derived. The fact that a trusts c in regard to completeness of q (sentence (8)) intuitively means that a trusts c for the fact that, if the train has arrived in the railway station then this information has been transmitted by c to a . By contraposition, this can be reformulated in the following way: if c has not transmitted this information to a then the train has not arrived. Then, in a situation where c has not transmitted the information, we can reason in that way: if the train would have been arrived, then c would have transmitted the information, therefore the train has not arrived.

More generally, the property of completeness is the dual of validity in the sense that, from the fact that an agent has not transmitted an information q , we can infer that q does not hold. Even if this kind of reasoning is not explicit in our mind, there are many situations where we apply it. For example, when a student observes that he is not in the list of students who have passed the examination, he infers that he failed, because he trusts the information source in regard to **completeness** of information in the list, not in regard to his validity. Even if some students who have not passed the examination are, wrongly, in the list, he can conclude that he failed because no student who passed the examination is missing.

The next derivation shows the kind of consequences that can be inferred from the fact that a trusts b in regard to his sincerity for q .

(1)	$I_{b,a} q$	in S
(2)	$K_a (I_{b,a} q)$	(1) and (OBS1)
(3)	$Tsinc_{a,b}(q)$	in S
(4)	$K_a (I_{b,a} q \rightarrow B_b q)$	(3) and definition of $Tsinc$
(5)	$K_a (B_b q)$	(2), (4) and (K2)

The first derivation has shown that from the fact that a trusts b for his validity for p , it can be inferred that a strongly believes p , while if a trusts b for his sincerity for q , but not for his credibility, it can only be inferred that a strongly believes that b believes q . More generally, we can easily prove that in our logic we have the following properties.

Properties

- $\vdash Tsinc_{a,b}(p) \rightarrow (I_{b,a}p \rightarrow K_a(B_b p))$
- $\vdash Tval_{a,b}(p) \rightarrow (I_{b,a}p \rightarrow K_a p)$
- $\vdash Tcoop_{a,b}(p) \rightarrow (\neg I_{b,a}p \rightarrow K_a(\neg B_b p))$
- $\vdash Tcomp_{a,b}(p) \rightarrow (\neg I_{b,a}p \rightarrow K_a(\neg p))$

These properties show what are consequences of performance or non-performance of information actions, depending on properties in regard to which an agent trusts an information source. In the absence of any kind of trust the only consequence of an information action on the receiver is that the action has been, or has not been, performed. That is what is stated by axiom schemas (OBS1) and (OBS2).

5. TOWARDS EXTENSION TO GRADED TRUST

In previous sections we have only considered situations where an agent a trusts or does not trust, an agent b in regard to some property. However, there are many practical applications where it might be convenient to express various levels of trust. In what follows we analyse a possible extension in this direction.

In our approach, instead of defining these levels in terms of some quantitative measure, we propose a purely qualitative definition. The idea is to characterise each trust level by a property which is attributed to some agent who is considered as a reference.

For instance, to define trust levels about meteorological expectations, we select from a set of information sources some who are used as references. Selected information sources are selected on the basis that there is a consensus about the ordering of their trust level with regard to some properties.

For instance, if we are interested in a proposition p about meteorological expectations in Toulouse, there might be a consensus about the fact that our trust about validity of Sud Radio channel is greater than our trust about validity of FR2 channel. That intuitively means that, even if it may happen that Sud Radio transmits information p while p is false, there is a conditional connection between the fact that Sud Radio has transmitted p and the fact that p is true, and this conditional connection is “stronger” than the conditional connection between the fact

that FR2 has transmitted p and the fact that p is true. In that case we say that the trust level about Sud Radio's validity is greater than trust level about FR2's validity. That could be supported by the fact that for local expectations, a local channel is better informed than a national channel.

If we consider the reference information sources: Sud Radio, FR2, TF1, ARTE and CNN, they can be used to define the partial order on trust levels represented in Figure 1 in terms of their validity with regard to p .

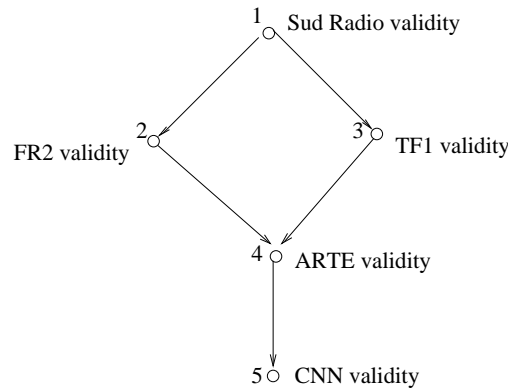


Figure 1.1 Trust levels defined by examples.

This partial order can be used to attribute trust levels to other information sources, and to reason about data transmitted by information sources that have the same level, or different levels. For instance, it might be assumed that the BBC trust level is the same as the ARTE trust level with regard to validity of p , and that the CANAL+ trust level is the same as the TF1 trust level.

Here, the information sources TF1 and ARTE play a similar role as numerical values, like 0.6 and 0.4, in the quantitative approach. We believe that the risk of misinterpretation by users who know TF1 and ARTE is reduced when they know that an information source has the same trust level as TF1 or ARTE, compared to the quantitative approach, where users know that an information source has a trust level equal to 0.6 or to 0.4.

From the definition of the partial order on trust levels, and assumptions about trust levels of information sources with regard to some properties (sincerity, credibility,...), and representation of a given situation in

terms of performed communication actions, we want to be able to infer consequences in terms of beliefs that are indexed by trust levels.

To extend our formal logic to deal with trust levels, the main idea is that in trust definitions given in previous sections, material implication is replaced by conditional connectives.

For instance, in the definition of trust about sincerity, sentence $I_{b,a}p \rightarrow B_b p$ is replaced by $I_{b,a}p \Rightarrow_i B_b p$. The intuitive meaning of connective \Rightarrow_i is defined in terms of reference information sources. Trust level i can be defined, for example, as the level at which we trust sincerity of a reference information source r_i with regard to sentence p , that is, by $I_{r_i,a}p \Rightarrow_i B_{r_i} p$.

The idea is that there is the same relation between the set of worlds where $I_{r_i,a}p$ is true and the set of worlds where $B_{r_i} p$ is true, as between the set of worlds where $I_{b,a}p$ is true and the set of worlds where $B_b p$ is true. We do not make explicit what this relation is, and we assume that in the context of a given application people have a rather clear idea of what it is. If we had a quantitative approach this relation might be, for instance, the value of conditional probability $Pr(B_{r_i}/I_{r_i,a})$, but, as we said before, we do not follow a quantitative approach.

In more formal terms, if we adopt Chellas's definition of conditionals (see Chellas, 1988 Chapter 10), a sentence of the form $\phi \Rightarrow_i \psi$ is true, iff the function $f_i(w, |\phi|)$ that interprets \Rightarrow_i assigns to a given world w and the truth set $|\phi|$ of ϕ , a set of worlds which is included in the truth set $|\psi|$ of ψ . If agent a strongly believes that $\phi \Rightarrow_i \psi$, that is, if we have $K_a(\phi \Rightarrow_i \psi)$, it is assumed that $f_i(w, |\phi|)$ is the same for every world w accessible by the accessibility relation that interprets K_a .

In general, trust levels are defined by sentences of the kind : $\phi(r_j, p) \Rightarrow_i \psi(r_j, p)$, where sentences ϕ and ψ may be of the form $I_{r_j,a}p$, or $B_{r_j} p$ or p . Trust levels are not in a one to one correspondence with reference information sources. For instance, sincerity of information source r_j can be used to define level i , and his credibility can be used to define level k .

To derive consequences of assumptions of the form $K_a(\phi \Rightarrow_i \psi)$ and $K_a \phi$, we consider as many doxastic modalities B_a^i as there are different trust levels. The reason is that we do not want to infer $K_a(\psi)$ but a weaker belief, represented by $B_a^i \psi$, which is indexed by trust level i . We accept the axiom schema:

$$(K_i) \quad K_a(\phi \Rightarrow_i \psi) \rightarrow (K_a \phi \rightarrow B_a^i \psi)$$

It is assumed that B_a^i modalities are normal modalities obeying (KD) schemas. To combine beliefs that have different trust levels we also accept the schemas:

If $i \geq j$, $B_a^i \phi \rightarrow B_a^j \phi$.

For every i , $K_a \phi \rightarrow B_a^i \phi$.

Notice that we do not have transitivity for conditional connectives. From $K_a(\phi \Rightarrow_i \psi)$ and $K_a(\psi \Rightarrow_i \theta)$ we cannot infer $K_a(\phi \Rightarrow_i \theta)$. Also, from $K_a \phi$ and $K_a(\phi \Rightarrow_i \psi)$ we can infer $B_a^i \psi$, but from $K_a \phi$, $K_a(\phi \Rightarrow_i \psi)$ and $K_a(\psi \Rightarrow_i \theta)$ we cannot infer $B_a^i \theta$. For that reason property of validity (resp. completeness) can no longer be defined from sincerity and credibility (resp. cooperativity and vigilance). They have to be independently defined.

If the fact that the trust level of agent a about property $prop$ of agent b with regard to p is i is denoted by $Tprop_{a,b}^i(p)$, we have:

$$Tsinc_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(I_{b,a}p \Rightarrow_i B_b p)$$

$$Tcred_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(B_b p \Rightarrow_i p)$$

$$Tcoop_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(B_b p \Rightarrow_i I_{b,a}p)$$

$$Tvig_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(p \Rightarrow_i B_b p)$$

$$Tval_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(I_{b,a}p \Rightarrow_i p)$$

$$Tcomp_{a,b}^i(p) \stackrel{\text{def}}{=} K_a(p \Rightarrow_i I_{b,a}p)$$

There are many derivations where, to derive consequences from properties of cooperativity or vigilance, we use the contraposition rule to infer $\neg q \rightarrow \neg p$ from $p \rightarrow q$. See, for instance, step (10) in the derivation presented in section 4. The problem with conditionals is that we do not have similar rule, since in general sentence: $(\phi \Rightarrow_i \psi) \rightarrow (\neg \psi \Rightarrow_i \neg \phi)$ is not valid.

Let us consider now the example presented in section 4 with the following representation of the situation.

$$S = \{Tsinc_{a,b}^1(p), Tsinc_{a,c}^2(\neg p), I_{b,a}(p), I_{c,a}(\neg p)\}$$

with the partial order on trust levels : $1 \geq 3$ and $2 \geq 3$.

Then we can draw the derivation:

- | | | |
|-----|--------------------|----------------|
| (1) | $I_{b,a}(p)$ | in S |
| (2) | $K_a(I_{b,a}p)$ | (1) and (OBS1) |
| (3) | $Tsinc_{a,b}^1(p)$ | in S |

(4)	$K_a(I_{b,a}(p) \Rightarrow_1 B_b(p))$	(3) and definition of $Tsinc^i$
(5)	$B_a^1(B_b(p))$	(2), (4) and (Ki)
(6)	$B_a^3(B_b(p))$	(5) and $1 \geq 3$
(7)	$I_{c,a}(\neg p)$	in S
(8)	$K_a(I_{c,a}(\neg p))$	(7) and (OBS1)
(9)	$Tsinc_{a,c}^2(\neg p)$	in S
(10)	$K_a(I_{c,a}(\neg p) \Rightarrow_2 B_c(\neg p))$	(9) and definition of $Tsinc^i$
(11)	$B_a^2(B_c(\neg p))$	(8), (10) and (Ki)
(12)	$B_a^3(B_c(\neg p))$	(11) and $2 \geq 3$
(13)	$B_a^3(B_b(p) \wedge B_c(\neg p))$	(6), (12) and properties of B^i

Notice that final conclusion is not inconsistent.

If we now add to S the assumptions: $\{Tval_{a,b}^1(p), Tval_{a,c}^2(\neg p)\}$. Instead of (5) and (11) we can infer $B_a^1(p)$ and $B_a^2(\neg p)$, and from the partial order we get the conclusion $B_a^3(p \wedge \neg p)$, which contradicts axiom schema (D). If assumptions $\{I_{b,a}(p), I_{c,a}(\neg p)\}$ are guaranteed to be true, this conclusion means that at least one of the assumptions in $\{Tval_{a,b}^1(p), Tval_{a,c}^2(\neg p)\}$ is false.

6. CONCLUSION

We have presented general concepts, and their formalisation in modal logic, to represent a specific view of trust which does not refer to normative aspects of interactions between agents but to epistemic aspects. We have introduced formal definitions for properties that characterise interactions: sincerity and credibility, and their dual counterpart: cooperativity and vigilance, which, as far as we know, have not been considered in the literature. We have also shown what are consequences of information actions on agent beliefs, depending on the properties attributed to trusted information sources. An original aspect in our work is to show what are consequences of **not performing** an information action.

As a matter of simplification we have defined very “fine grained” properties which refer to one specific sentence p . In fact, in many practical situations trust is not defined at such a detailed level. However, it would not be a difficult work to extend definitions of these properties to situations where they refer to a set of sentences of a given sort, for instance, to all the sentences of the form $p(x,y)$, as we did in (see Demolombe, 1996; Demolombe, 1997 for a formal definition of topics). Also, definitions could be extended to all the sentences that are about a given topic Demolombe and Jones, 1999. For instance, an agent might be trusted for his sincerity about data he communicates that are about health, but not for data that are about income.

Further works based on the different kinds of trust presented here might be about another approach of speech acts (Cohen and Perrault, 1979; Jones, 1990; Cohen and Levesque, 1990). Another possible direction is to investigate how these definitions of trust can be used to analyse deceptions.

Acknowledgments

We would like to thank Andrew J.I. Jones for his fruitfull comments about this paper.

Notes

1. Here we take inspiration from a talk given by Andrew J.I. Jones in an informal workshop organised at the University of Lisbon in September 1997. One of his intuitive ideas is that “agent a trusts agent b if a believes that b will act in accordance with a norm which a believes b accepts”. More details can be found in A.J.I. Jones, 1983.

References

- A.J.I. Jones (1983). *Communication and meaning: An essay in applied modal logic*. Synthese Library. Reidel.
- Chellas, B. F. (1988). *Modal Logic: An introduction*. Cambridge University Press.
- Cohen, P. and Levesque, H. (1990). Rational interaction as the basis for communication. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communications*. The MIT Press.
- Cohen, P. R. and Perrault, C. R. (1979). Elements of a Plan-Based Theory of Speech Acts. *Cognitive Science*, 1:177–212.
- Demolombe, R. (1996). Validity Queries and Completeness Queries. In *Proc. of 9th International Symposium on Methodologies for Intelligent Systems*.
- Demolombe, R. (1997). Answering queries about validity and completeness of data: from modal logic to relational algebra. In Andreassen, T., Christiansen, H., and Larsen, H. L., editors, *Flexible Query Answering Systems*. Kluwer Academic Publishers.
- Demolombe, R. and Jones, A. (1999). On sentences of the kind “sentence “p” is about topic “t”: some steps toward a formal-logical analysis. In H-J. Ohlbach and U. Reyle, editor, *Logic, Language and Reasoning. Essays in Honor of Dov Gabbay*. Kluwer Academic Press.
- Jones, A.J.I. (1990). Toward a Formal Theory of Communication and Speech Acts. In Cohen, P., Morgan, J., and Pollack, M., editors, *Intentions in Communications*. The MIT Press.

- Lerch, F. and Prietula, M. (1989). How do we trust machine advice. In *Third International Conference on Human-Computer Interaction*.
- March, S. (1994). Trust in distributed Artificial Intelligence. *Lectures Notes in Computer Science*, 830.
- P.Oerbaek (1995). Can you Trust your data. *Lectures Notes in Computer Science*, 915.
- Thompson, K. (1984). Reflections on Trusting Trust. *Communications of ACM*, 27(18).