Trust and Agency in the context of Communication

Robert Demolombe[1]

Institut de Recherche en Informatique de Toulouse
France
`robert.demolombe@orange.fr`

Abstract. The communication process is analyzed on the basis of the notions of trust and agency. The aim of the paper is to clarify the role played by causality on one hand and the role played by logical consequences of assumptions about trust in information sources on the other hand. The first part is informal and the second part refers to the logical framework of modal logic though it requires a quite limited background in this area.

1 Introduction

The communication process may involve two different concepts: agency and trust. Agency explains some causal consequences of an informing action and trust explains some logical consequences about what an agent believes. However, it is not easy to make a clear distinction between these two kinds of consequences because they may be imbricated.

Let's consider, for instance, a situation in an airport where John is waiting the arrival of the flight AF001. In the arrival hall there is a screen where it is indicated that this flight should arrive at 11.30. At 11.20 a speaker announces that the flight is delayed of 15 minutes. It may be that John has heard this announcement. In that case John believes that the speaker has said: "*the flight number AF001 is delayed of fifteen minutes*". It may also be that John believes that the speaker has said: "*the flight number AF001 is delayed of fifty minutes*" and it may also be that at this moment John was talking with some friend and that he has heard nothing.

We can also imagine another similar situation where at 11.20 there is the text "*DELAYED 15 mn*" that appears in the screen. If John is looking at the screen he believes that someone in the airport has run a program which has modified the line of the flight AF001 on the screen.

In both cases, oral announcement or screen modification, it may be assumed that John believes that the meaning of the transmitted message is that the flight should arrive at 11.45.

If, in addition, it is assumed that John trusts the person who has performed the informing action[1] in his reliability, it can logically be inferred that John believes that the flight shall arrive at 11.45.

If, in a similar situation, at 11.20 or at some further moment no informing action about the fact that this flight is delayed has been performed, and John trusts the information source who is in charge to announce a flight delay in the property of announcing a flight delay if there is a flight delay, then John believes that the flight AF001 is not delayed. It is worth noting that in that case what is derived by John is the logical consequence of the fact that an informing action has not been performed.

The aim of this paper is to clarify the roles played by causal effects of informing actions and by logical consequences of trust in the communication process.

Since there are several variants in the definitions of causality and agency and in the definitions in trust, in the first part of the paper are given semi-formal definitions of these concepts. In the second part are presented in a more formal way how these concepts interacts in the communication process. Then, it will be shown that we could accept a variant of the notion of agency which takes into account the different kinds of trust.

Most of the content of the paper is intended to be read by people who have a limited familiarity with formal logic. Then, people who have this background could find that some formal details are missing.

2 Informal section

2.1 Causality and Agency

Surveys about the most significant approaches of causality can be found in [21, 1]. In this section are briefly presented some of the most significant definitions of this concept. We have slightly changed authors' notations for the sake of homogeneity of the presentation.

G.H. von Wright in his paper [22] defines causality with respect to three states of the world: w the state of the world where the agent starts to act, w' the state of the world where the agent ends to act and w'' the state of the world where we would be if the agent had not been active, that is where the only changes are due to the laws of nature. In the following this world is called the "*counterfactual world*" (see [17]).

Then, the action performed by an agent has caused that a property represented by p is obtained iff p is not the case in w, p is the case in w' and p is not the case in w''. In that situation we say that in w' the agent has brought it about that p. See figure 1.

For instance, if p means that the door is open, the agent has caused that the door is open iff in w the door is not open, in w' the door is open and in w'' the door is not open.

---

[1] In most cases this kind of trust is implicitly adopted by the users. Nevertheless, it is necessary to draw the following consequences.
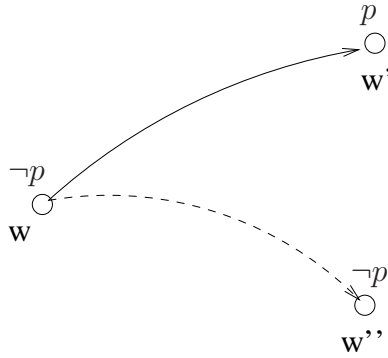
Fig. 1. To bring it about that p.

In von Wright definitions it is assumed that agent's action leads to a unique state of the world. That means that what is called an action should be interpreted has a specific occurrence of an action.

In another definition proposed by I. Pörn in [19] an agent brings it about that p iff the action he does necessarily leads to a situation where p holds and it might be that p does not hold if this action is not performed in a situation which is the same as the current situation, except that this particular action is not performed. This second condition is called the "*counteraction condition*". Here the notion of necessitation is defined with respect to the set of hypothetical situations where the agent does at least the action he does in the actual situation.

A significant difference with regard to von Wright's definition is that Pörn does not make explicit reference to the world where the action starts and the world where the action ends.

R. Hilpinen in [11] has proposed a definition of agency which has similarities with von Wright's definition since he also considers three worlds, and similarities with Pörn since his definition is based on some notion of necessity. Nevertheless, the Hilpinen's necessity condition refers to the set of all the possible performances of a given action type a instead of the set of situations where the agent may do any actions in addition to the action he actually does.

For example, Hilpinen might consider the set of possible instances of the action type "to close the door", while Pörn might consider the set of situations where the agent closes the door and also smokes a cigarette or closes the door and also speaks to someone else ... etc ... The counteraction condition given by Hilpinen is more precise. It is defined with respect to the performance of an instance of an action type.

The definition of agency we have adopted in this paper takes inspiration from the three authors we have presented before. It is based on the definition

presented in [10]. Instead of the notion of "*action*" it is based on the notion of "*act*"[2].

An act is a pair made of an actor i (an agent) and an action type a. This pair is denoted by $i : a$. An actor can perform an action type in many different ways and a relation is defined between three worlds: the world w where the action starts, the world w' where a given instance of the action type has been performed and the world w'' which is the counterfactual world of this particular instance of the action type.

It may be that in w' other acts than $i : a$ have been performed in parallel. The set of all the acts performed in w' (including the acts performed by other agents than i) is called *all*. Then, the set of acts performed in w'' is denoted by: $all - \{i : a\}$.

In this definition we say that the act *act* is going to bring it about that p in w when the set of acts performed in parallel is *all* iff

1. for all the worlds w' where the set of acts *all* ends it is the case that p holds, and
2. there exists a world w'' where only $all - \{i : a\}$ are performed and where p does not hold.

2.2 Trust

There are many definitions of trust (see [2]). However, most of them agree on the fact that trust is a form of the truster's belief in some trustee's property. In [5, 6, 18] we have proposed several kinds of trustee's properties which are relevant in the context of communication. They are first presented by examples.

Let's consider a situation where Peter is a consultant in the field of finance and John is looking for advices about buying or selling stocks of the company XYZ. In particular John is interested in the truth value of the proposition p: "the value of the stock of company XYZ is going to strongly decrease".

The properties of the information source Peter are defined in terms of the relations between the fact that Peter has informed John about p, the fact that Peter believes that p is true and the the fact that p is true. These relations have the form of conditionals. For example Peter's sincerity means that IF Peter informs John about the fact that p is true, THEN Peter believes that p is true. Then, John trusts Peter in his sincerity about p means that John believes that Peter is sincere with regard to John about p.

The different kinds of properties of an information source are defined in this example as follows.

Sincerity. Peter is sincere with regard to John about p iff IF Peter informs John about the fact that p is true, THEN Peter believes that p is true.

Competence. Peter is competent about p iff IF Peter believes that p is true, THEN it is the case that p is true.

---

[2] In [10] agency is defined for joint acts that cause some effects. Here, for simplification, the definition is restricted to the effects which are caused by a unique act.

Validity. Peter is valid with regard to John about p iff IF Peter informs John about the fact that p is true, THEN it is the case that p is true.

It is worth noting that these properties are not logically independent. Indeed, validity can be derived from sincerity and competence.

From each property of the form "IF A, THEN B" we can define a dual property of the form "IF B, THEN A". That leads to the following properties.

Cooperativity. Peter is cooperative with regard to John about p iff IF Peter believes that p is true, THEN Peter informs John about the fact that p is true.

Vigilance. Peter is vigilant about p iff IF it is the case that p is true, THEN Peter believes that p is true.

Completeness. Peter is complete with regard to John about p iff IF it is the case that p is true, THEN Peter informs John about the fact that p is true.

These properties can be generalized to any truster i, any trustee j and any proposition p [3].

They are represented in the figure 2 where the edges can be seen as implication operators and the labels of the edges as the name of the properties.
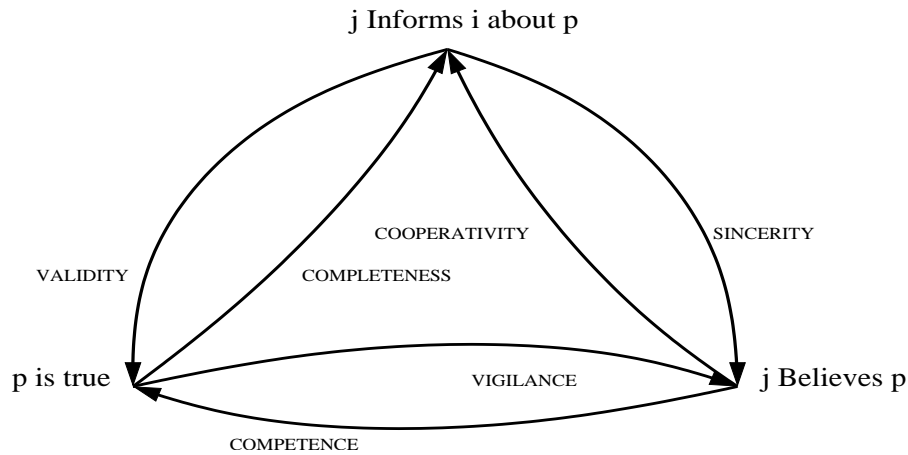
Fig. 2. Relationships between believing, informing and truth.

2.3   Communication

In this section we analyze the communication process on the basis of the notions of agency and trust we informally adopted before.

---

[3] A. J. I. Jones in [14–16] has defined the properties of reliability and sincerity which respectively correspond to what we call validity and sincerity. However, the other properties like competence or completeness (the dual of validity) are not mentioned in these works.

The first step in this process is the performance of an informing action at the physical level. As we have seen in the above examples at this level an agent j does some physical action the effect of which is that a message (we could also say a "signal") is produced. It may be a sound in the case of an oral announcement in an airport or it may be a text that appears on a screen or any kind of physical support. This physical message is denoted by "p".

At the end of this step, on the basis of the definition of agency we have adopted, we can clearly check whether it is the case that some effect has been caused by an action performed by agent j.

In a second step, it may be that an another agent i has perceived the message "p". To be more precise we must say that i believes that the message "p" has been produced by j. Indeed, it may be, if the communication process does not work well, that i believes that "p" has been produced while it is another message "q" which has been produced.

It can be assumed in a third step that agent i assigns some meaning to the message "p". This meaning is that a proposition denoted by p is true. Then, on the basis of this assumption, the logical consequence of what i has perceived is that i believes that j has performed an informing action the effect of which is that a message has been produced and that this message means that p is true[4].

Can we say that what i believes at this step has been caused by the informing action? To answer this question we have to check whether i already had this belief or not before j performs the informing action.

If it is assumed, in addition, that, for example, i trusts j in is validity about p and that i has some reasoning capacities, then it can be inferred that i believes that p is true.

At this level we can ask a similar question: does the fact that i believes that p is true has been caused by the informing action performed by j? The answer is "*yes*" if it is the case that before performance of this action i did not believe that p is true and i would not believe that p is true in a counterfactual situation.

Let's consider now another scenario where i trusts j in his completeness about p. From the definition of completeness, if we can apply the contraposition rule to the IF A, THEN B proposition, we can infer IF ¬B, THEN ¬A. Then, from i's trust in completeness we can infer that i believes that IF j does not inform i about p, THEN it is not the case that p is true.

The same kind of reasoning which has been shown for trust in validity could lead to the conclusion that if j has not informed i about p, then i believes that it is not the case that p is true. However, there is a significant difference which can be understood if we consider the two concrete examples we have seen above.

Indeed, in the case of the arrival time of the flight AF001, if John trusts the airport speaker about his completeness he cannot infer that the flight is not delayed at any time before 11.30. For instance, if no announcement has been done

---

[4] The distinction between the action of sending a message and the informing action is similar as the distinction between a locutionary act and an illocutionary act in the J. Searle's Speech Act Theory [20]. See also J. I. I. Jones in [14, 15].

at 10.00 that does not guarantee that a delay announcement will not happen further, for instance at 11.15.

In the example of the finance consulting it is the same. Even if John trusts Peter in his completeness about p he can infer that the value of the stock of XYZ is not going to decrease only in some well defined circumstances.

We can imagine a third example where these circumstances are easier to define. Let's consider a speaker who is announcing the list of students who have succeeded at a given exam. If John is such a student and he trusts the speaker in his completeness about the set of students who have succeeded, after the speaker has ended to announce the list, if the name of John has not been announced, John infers that he has not succeeded. Because he believes that the list was complete.

These examples show that the properties assigned by the truster to the trustee are properties that holds only if some specific conditions about the context are satisfied. In semi formal terms, these properties must have the form: IF C, THEN (IF A, THEN B), where C his a proposition that characterizes this context.

3 Formal section

To represent more formally the notions of agency and trust and the derivations involved in the communication process we need first to briefly introduce the formal language.

Language: The language is a First Order Modal Language (see [3]). It is defined as usual and its modal operators are: $Bel_i(p)$, $Br^+_{all,i:a}(p)$, $Inform_{i,j}(p)$.

The intuitive meaning of these operators is:

$Bel_i(p)$: agent i believes that the proposition represented by p is true.

$Br^+_{all,i:a}(p)$: agent i is going to bring it about that p by doing an action of type a in a context where the set of acts performed in parallel is all[5].

$Nec^+_{all,i:a}(p)$: p is true after any instance of the set of acts all has been performed (notice that all includes the act $i : a$).

The operator $Nec^+_{all,i:a}(p)$ is intended to represent the fact that a proposition p is true after performance of a set of acts even if its truth has not been caused by these acts. That is the case for instance for laws of nature (for example, the proposition: *time is flying* holds whatever the agents do).

$Inform_{i,j}(p)$: agent i has transmitted to agent j a message the meaning of which is that a proposition represented by p is true.

Axiomatics. The axiomatics of these modalities is briefly defined below.

Bel_i is a normal modality which obeys the system (KD).

$Br^+_{all,i:a}$: is a classical modality which only obeys the following inference rule.

(REB) If $\vdash \phi \leftrightarrow \psi$, then $\vdash Br^+_{all,i:a}(\phi) \leftrightarrow Br^+_{all,i:a}(\psi)$

$Inform_{i,j}$: is a classical modality which only obeys the inference rule:

---

[5] In $Br^+_{all,i:a}(p)$ the + is used to make the distinction with $Br_{all,i:a}(p)$ which means that agent i has brought it about that p. $Br_{all,i:a}(p)$ can be defined from $Br^+_{all,i:a}(p)$.

(REI) If $\vdash \phi \leftrightarrow \psi$, then $\vdash Inform_{i,j}(\phi) \leftrightarrow Inform_{i,j}(\psi)$

Frame.

A frame is a tuple which has the following components:

- A domain D for the interpretation of the terms of the language.
- A set of possible worlds W.
- A set of relations R_i defined on $W \times W$ for the interpretation of the operators Bel_i.
- A set of relations $R_{all,i:a}$ defined on $W \times W \times W$ for the interpretation of the operators $Br^+_{all,i:a}$.
- A set of functions $I_{i,j}$ defined from W to W^W for the interpretation of the operators $Inform_{i,j}$.

The intuitive meaning of $R_{all,i:a}(w, w', w'')$ is that the set of acts *all* which have started in w have ended in w' and w'' is a counterfactual world of w' where all the acts in *all* have been performed except the act $i : a$.

Model. A model M is a frame which validates the axiom schemas and inference rules of the above operators and of Classical First Order Logic. The fact that a formula ϕ is true at the world w of the model M is denoted as usual by: $M, w \models \phi$ and the fact that ϕ is a valid formula is denoted by: $\models \phi$.

Satisfiability. The satisfiability conditions are defined below.

We have $M, w \models Bel_i(\phi)$ iff for every world w' such that $R_i(w, w')$ we have $M, w' \models \phi$.

We have $M, w \models Br^+_{all,i:a}(\phi)$ iff

1) we have $M, w \models \neg \phi$ and

2) for every world w' and for every world w'' such that $R_{all,i:a}(w, w', w'')$ we have $M, w' \models \phi$ and

3) there exist a world w'_1 and a world w''_1 such that $R_{all,i:a}(w, w'_1, w''_1)$ and $M, w''_1 \models \neg \phi$.

The satisfiability conditions of the operator $Br^+_{all,i:a}$ assigns a more precise meaning to the notion of agency.

We have $M, w \models Nec^+_{all,i:a}(\phi)$ iff

for every world w' and for every world w'' such that $R_{all,i:a}(w, w', w'')$ we have $M, w' \models \phi$.

We have $M, w \models Inform_{i,j}(\phi)$ iff the set of worlds $I_{i,j}(w)$ is the set of worlds w' such that $M, w' \models \phi$.

Now, we can give formal definitions to the different kinds of trust.

Trust in sincerity.

$TrustSinc_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \rightarrow (Inform_{j,i}(p) \rightarrow Bel_j(p)))$

Trust in competence.

$TrustComp_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \rightarrow (Bel_j(p) \rightarrow p))$

Trust in validity.

$TrustVal_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \rightarrow (Inform_{j,i}(p) \rightarrow p))$

Trust in cooperativity.

$TrustCoop_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \rightarrow (Bel_j(p) \rightarrow Inform_{j,i}(p)))$

Trust in vigilance.

$$TrustVig_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \to (p \to Bel_j(p)))$$

Trust in completeness.

$$TrustComplt_{i,j}(p) \stackrel{\text{def}}{=} Bel_i(ctxt \to (p \to Inform_{j,i}(p)))$$

ctxt is a proposition which restricts the circumstances where the truster trusts the trustee.

Now, we can formally define some terms and predicates which are relevant in the context of communication.

Informing act. $i : inf_{j,"p"}$ is an informing act where i does an action of the type: $inf_{j,"p"}$. The meaning of this action type is to transmit to j the signal $"p"$.

Predicate $Transmitted$**.** The predicate $Transmitted_{i,j}("p")$ means that i has transmitted to j the message $"p"$ ($"p"$ is a constant).

We present now a derivation which formally shows the derivation process. This derivation shows the point of view of a logician who is reasoning about what the agents do and believe.

Let's assume first that in the world w agent j is going to bring it about that $Transmitted_{j,i}("p")$ is true by doing an action of the type $inf_{i,"p"}$. Then, we have:

(1) $M, w \models Br^+_{all,j:inf_{i,"p"}}(Transmitted_{j,i}("p"))$

From (1) and the satisfiability conditions of the operator $Br^+_{all,j:inf_{i,"p"}}$, for every world w' such that $R_{all,j:inf_{i,"p"}}(w, w', w'')$ we have:

(2) $M, w' \models Transmitted_{j,i}("p")$

Let's assume now that in the world w agent i pays attention to the fact that j transmits $"p"$ to him. This assumption is formally represented by:

(3) $M, w \models Nec^+_{all,j:inf_{i,"p"}}(Transmitted_{j,i}("p") \to Bel_i(Transmitted_{j,i}("p")))$

From (3) and the satisfiability conditions of the operator $Nec^+_{all,j:inf_{i,"p"}}$ in w' we have:

(4) $M, w' \models Transmitted_{j,i}("p") \to Bel_i(Transmitted_{j,i}("p"))$

From (2) and (4) we have:

(5) $M, w' \models Bel_i(Transmitted_{i,i}("p"))$

Let's assume that in w, when agent j has performed any instance of the act $j : inf_{i,"p"}$, agent i assigns to the message $"p"$ the meaning that p is true. Then, we have:

(6) $M, w \models Nec^+_{all,j:inf_{i,"p"}}(Bel_i(Transmitted_{j,i}("p") \to Inform_{j,i}(p)))$

From the satisfiability conditions of the operator $Nec^+_{all,j:inf_{i,"p"}}$, in w' we have:

(7) $M, w' \models Bel_i(Transmitted_{j,i}("p") \to Inform_{j,i}(p))$

From (5) and (7) we have:

(8) $M, w' \models Bel_i(Inform_{j,i}(p))$

Let's assume now that in w agent i trusts agent j in his validity about p when agent j has performed any instance of the act $j : inf_{i,"p"}$. Then, we have:

(9) $M, w \models Nec^+_{all,j:inf_{i,"p"}}(TrustVal_{i,j}(p))$

From the satisfiability conditions of the operator $Nec^+_{all,j:inf_{i,"p"}}$ in w' we have:

(10) $M, w' \models TrustVal_{i,j}(p)$

Then, from $TrustVal$ definition we have:

(11) $M, w' \models Bel_i(ctxt \to (Inform_{j,i}(p) \to p))$

From (8) and (11) we have:

(12) $M, w' \models Bel_i(ctxt \to p)$

Let's assume that in all the worlds like w' i believes that proposition $ctxt$ holds. Then, we have:

(13) $M, w' \models Bel_i(ctxt)$

From (12) and (13) we have:

(14) $M, w' \models Bel_i(p)$

The fact that we have (14) in every world w' where j has performed the act $j : inf_{i,"p"}$ is not enough to conclude that in w j is going to bring about that i believes that p is true. We also need to satisfy the conditions 1) and 3) in the satisfiability conditions of the operator Br^+.
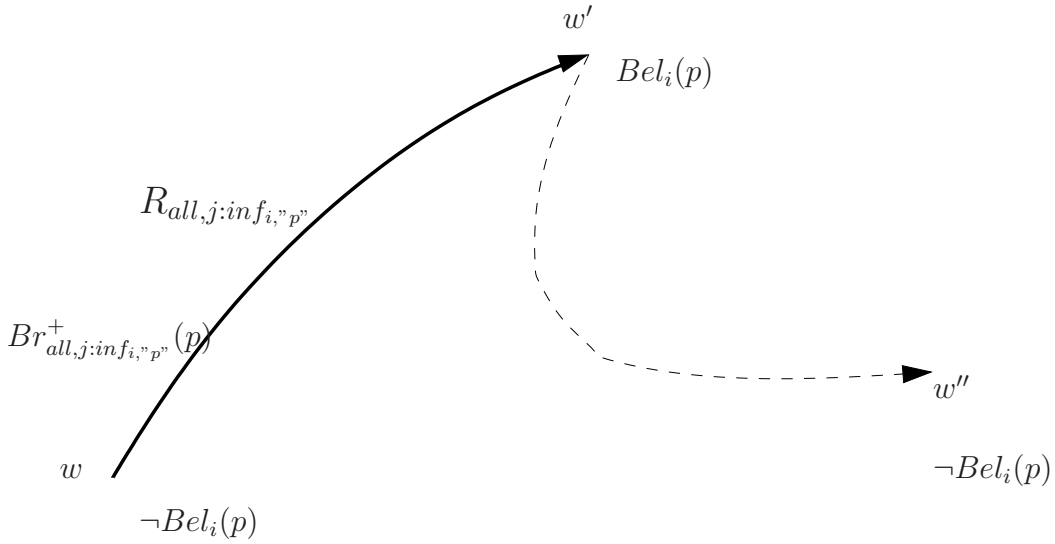
The condition 1) requires to have:

(15) $M, w \models \neg Bel_i(p)$

The condition 3) requires that there exist a world w'_1 and a world w''_1 such that we have $R_{all,i:a}(w, w'_1, w''_1)$ and:

(16) $M, w''_1 \models \neg Bel_i(p)$

If (15) and (16) hold, and the assumptions that allow to infer (14) hold, we have:

(17) $M, w \models Br^+_{all,j:inf_{i,"p"}}(Bel_i(p))$ (see figure 3).

w'

$Bel_i(p)$

$R_{all,j:inf_{i,"p"}}$

$Br^+_{all,j:inf_{i,"p"}}(p)$

w''

w

$\neg Bel_i(p)$

$\neg Bel_i(p)$

Fig. 3. Agent j is going to bring it about that agent i believes p.

(17) intuitively means that in the world w agent j is going to bring it about that agent i believes that p is true by doing the act $j : inf_{i,"p"}$, in a context where the set of performed acts is all.

This derivation exhibits the assumptions that allow to infer (17). These assumptions are: (3) the message sent by j is received by i

(6) i assigns the meaning p to this message

(9) i trusts j in his validity about p

(13) i believes that he is in a context where he can trust j

(15) before j sends the message i does not believe p

(16) if j had not send the message, i would not believe p; that is, the other acts in all do not inform i about p.

A similar kind of derivation allows to characterize the assumptions that lead to the conclusion that j is going to bring it about that agent i believes that j believes that p is true, if i trusts j in his sincerity.

The situation where in the world w agent j has not informed agent i about p cannot be analyzed on the basis of a trivial duality between trust in validity and trust in completeness.

Let's assume that in w agent i trusts j in his completeness about p. We have:

(1') $M, w \models TrustComp_{i,j}(p)$

From the definition of trust in completeness we have:

(2') $M, w \models Bel_i(ctxt \rightarrow (p \rightarrow Inform_{j,i}(p)))$

Let's assume that i believes that in w the context $ctxt$ holds. We have:

(3') $M, w \models Bel_i(ctxt)$

From(2') and (3') we have:

(4') $M, w \models Bel_i(p \rightarrow Inform_{j,i}(p))$

By contraposition, from (4') we have:

(5') $M, w \models Bel_i(\neg Inform_{j,i}(p) \rightarrow \neg p)$

Let's assume that in w i believes that j has not informed i. Then, we have:

(6') $M, w \models Bel_i(\neg Inform_{j,i}(p))$

From (5') and (6') we have:

(7') $M, w \models Bel_i(\neg p)$

(7') is a logical consequence of the above assumptions. Now, we can ask the question: "*is it the case that agent j has brought it about that i believes that p is false in the world w*"?

Answering this question deserves further researches, but we are inclined to think that the answer is "no". Our argument to support this answer is that if (6') holds, that is, if i believes that j has not informed i in w, then, it does not exist a world w_0 and a world w_0'' such that we have: $R_{all,j:inf_{i,"p"}}(w_0, w, w_0'')$. Since the notion of agency has been defined for an action which has been performed we cannot infer that (7') has been caused by some informing action performed by j.

The definition of agency is usually defined as a relationship between an action and a proposition that holds after the action been preformed. However, one could say that we just need to modify the definition of agency in order to accept statements of the kind: "*i believes that p is false because it is not the case that i*

has been informed by some agent who is trusted by i in his completeness about p". Nevertheless, in this approach we would need to define a counteraction world with respect to an action which has not been performed. That is certainly a non trivial issue.

4 Conclusion

Formal logic has been used to give clear definitions of several kinds of trust in information sources and to help to understand the roles played by causality and agency in the process of information communication. It has been shown that if some assumptions are accepted communication can be interpreted as a causal process in the case where an informing action has been performed. In the cases where the informing action has not been performed the question of viewing the communication as a causal process is an open question.

There are still many other open issues to be investigated. We just mention some of them below.

The definition of agency is more complex when several agents are acting together (see [10]) and the extension of the notion of trust to a group of agent should deserves further researches. Also, the notion of trust must be refined when several agents are involved in the communication process. That raises the issue of trust transitivity (see [9]).

The notions of causality and agency could be refined depending on the type of agents we are interested in. They may be human agents, software agents or institutional agents (like a bank) (see [8]). The basic differences are that in the case of software agents the causal relationship is only determined by the laws of nature, in the case of human agents the notion of free choice plays a significant role (see [12, 13]) and in the case of institutional agent causality is indirectly determined by institutional norms which define who are the human agent who are acting on the behalf of an institution (see [4]).

Another non trivial issue is to define the notion of agency in the case where we move to graded trust which is defined for several qualitative levels like in [7].

Acknowledgements. We would like to thank a reviewer who has pointed out some mistakes.

References

1. L. Aqvist. Old foundations for the logic of agency and action. *Studia Logica*, 72, 2002.
2. C. Castelfranchi and R. Falcone. *Trust Theory: A Socio-Cognitive and Computational Model*. Wiley, 2010.
3. B. F. Chellas. *Modal Logic: An introduction*. Cambridge University Press, 1988.
4. M. Colombetti and M. Verdicchio. An analysis of agent speech acts as institutional actions. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the first international joint conference on Autonomous Agents and Multiagent Systems*, pages 1157–1166. ACM Press, 2002.

5. R. Demolombe. To trust information sources: a proposal for a modal logical framework. In C. Castelfranchi and Y-H. Tan, editors, *Trust and Deception in Virtual Societies*. Kluwer Academic Publisher, 2001.

6. R. Demolombe. Reasoning about trust: a formal logical framework. In C. Jensen, S. Poslad, and T. Dimitrakos, editors, *Trust management: Second International Conference iTrust (LNCS 2995)*. Springer Verlag, 2004.

7. R. Demolombe. Graded Trust. In R. Falcone and S. Barber and J. Sabater-Mir and M.Singh, editor, *Proceedings of the Trust in Agent Societies Workshop at AAMAS 2009*, 2009.

8. R. Demolombe. Relationships between obligations and actions in the context of institutional agents, human agents or software agents. *Journal of Artificial Intelligence and Law*, 19(2), 2011.

9. R. Demolombe. Transitivity and propagation of trust in information sources. An Analysis in Modal Logic. In J. Leite and P. Torroni and T. Agotnes and L. van der Torre, editor, *Computational Logic in Multi-Agent Systems, LNAI 6814*. Springer Verlag, 2011.

10. R. Demolombe. Causality in the context of multiple agents. In T. Agotnes and J. Broersen and D. Elgesem, editor, *Deontic Logic in Computer Science, LNAI 7393*. Springer Verlag, 2012.

11. R. Hilpinen. On Action and Agency. In E. Ejerhed and S. Lindstrom, editors, *Logic, Action and Cognition: Essays in Philosophical Logic*. Kluwer, 1997.

12. J. Horty. *Agency and deontic logic*. Oxford University Press, 2001.

13. J.F. Horty and N. Belnap. The deliberative STIT: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.

14. A.J.I. Jones. *Communication and meaning: An essay in appplied modal logic*. Synthese Library. Reidel, 1983.

15. A.J.I. Jones. Toward a Formal Theory of Communication and Speech Acts. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communications*. The MIT Press, 1990.

16. A.J.I. Jones. On the concept of trust. *Decision Support Systems*, 33, 2002.

17. D. Lewis. *Counterfactuals*. Harvard University Press, 1973.

18. E. Lorini and R. Demolombe. From trust in information sources to trust in communication systems: an analysis in modal logic. In J. Broersen and J.-J. Meyer, editors, *International Workshop on Knowledge Representation for Agents and Multi-agent Systems (KRAMAS 2008): Postproceedings*, LNAI. Springer-Verlag, 2009.

19. I. Pörn. Action Theory and Social Science. Some Formal Models. *Synthese Library*, 120, 1977.

20. J. R. Searle. *Speech Acts: An essay in the philosophy of language*. Cambridge University Press, New-York, 1969.

21. K. Segerberg. Getting started: beginnings in the logic of action. *Studia Logica*, 51:347–378, 1992.

22. G. H. von Wright. *Norm and Action*. Routledge and Kegan, 1963.