Distrust is not Always the Complement of Trust (Position Paper)

Célia da Costa Pereira
Università degli Studi di Milano, DTI
{pereira}@dti.unimi.it

Abstract. We believe that *distrust* can be as important as *trust* when agents are making a decision. An agent may not trust a source because of lack of positive evidence, but this does not necessarily mean the agent distrusts the source. Trust and distrust have to be considered as two separate concepts which can coexist. We are aware that an adequate way to take this fact into account is by considering explicitly not only the agent's degree of trust in a source but also its *independent* degree of distrust. Explicitly taking distrust into account allows us to mark a clear difference between the distinct notions of *negative* trust and *insufficient* trust. More precisely, it is possible, unlike in approaches where only trust is explicitly accounted for, to "weigh" differently information from *helpful*, *malicious*, *unknown*, or *neutral* sources.

1 Introduction and Motivations

Interaction is fundamental in a multi-agent system, allowing agents to cooperate to achieve their goals. When interchanging information, the extent to which a rational agent changes its beliefs may depend on several factors, like, for example, the trustworthiness of the agent providing new information, the agent's attitude towards information coming from unknown agents, or agents the agent knows as being malicious, or agents the agent knows as providers of usually correct information, and so on.

The main lack in most existing works on trust or using trust is the way the concept of *distrust* is implicitly considered, that is, as the complement of *trust* (*trust* $=$ $1 -$ *distrust*). However, things are not always so simple. Trust and distrust may derive from different kinds of information (or from different sides of the personality) and, therefore, can coexist without being complementary [4, 9, 15]. For instance, one may not trust a source because of lack of positive evidence, but this does not necessarily mean (s)he distrusts it. Taking distrust explicitly into account allows an agent, e.g., to avoid dropping a goal just because favorable information comes from an unknown agent (neither trusted nor distrusted) — the absence of trust does not always mean full distrust.

We believe that an adequate way to take these facts into account is by considering explicitly not only the agent's trust degree in other agents but also its *independent* degree of distrust. The trustworthiness of a source can be represented as a (trust, distrust) pair, and intuitionistic fuzzy logic [1] can be used to represent the uncertainty on the trust degree introduced by the explicit presence of distrust.

Dagstuhl Seminar Proceedings 09121
Normative Multi-Agent Systems
http://drops.dagstuhl.de/opus/volltexte/2009/1915

Example John thinks his house has become too small for his growing family and would like to buy a larger one. Of course, John wants to spend as little money as possible. A friend who works in the real estate industry tells John prices are poised to go down. Then John reads in a newspaper that the real estate market is weak and prices are expected to go down. Therefore, John's desire is to wait for prices to lower before buying. However, John later meets a real estate agent who has an interesting house on sale, and the agent tells him to hurry up, because prices are soaring. On the way home, John hears a guy on the bus saying his cousin told him prices of real estate are going up.

We will see below which degrees could be assign by John to these different sources of information. Note that here, although numerical, the source degrees have only an ordinal significance.

2 Related Work: From Trust to Distrust

2.1 From Trust ...

The term *trust* has a variety of meanings in the literature [14]. Demolombe [5] defines trust as a mental attitude of an agent with respect to another agent. He considers the agent's attitudes as a sort of belief about some property about another agent. He started by proposing a definition of trust as a binary concept (an agent trusts or does not trust another agent); then he introduced the notion of *graded trust* which is more suited to representing real situations where trust is perceived as less rigid. He also proposed a formal defintion for both *trust with respect to topics* and *conditional trust*.

Definition 1 (trust with respect to topics)
The fact that an agent a trust the agent b with respect to $prop$ for the topic t, $Tprop_{a,b}(t)$, is defined as follow:

$$Tprop_{a,b}(t) \equiv \forall \text{“}p''(A(t, \text{“}p'') \rightarrow Tprop_{a,b}(p),$$

where $A(t, \text{“}p'')$ means that the sentence named by “p'' is about the topic t.

The justification for this first proposal is that in general an agent trusts another agent for all the propositions related to a given topic.

Definition 2 (Conditional Trust)
The fact that an agent a trust the agent b for p in the circumstances represented by q, $Tprop_{a,b}(p|q)$, is defined as follow:

$$Tprop_{a,b}(p|q) \equiv K_a(q \rightarrow prop(p)),$$

where $prop(p)$ is any property about p.

The justification for this second proposal is that, in real life, there are many situations where an agent trusts another only in some particular circumstances. This is called for example, *context-awareness trust* [17] or *decision trust* [11].

Castelfranchi and Falcone [2] claimed that trust is much more than a subjective probability and pointed out to the necessity for a cognitive view of trust as a complex

structure of beliefs (and goals) determining both a "degree of trust" (instead of a simple probability factor) and an estimation of risk. However, they expressed their awareness that in some situations, it should be important to consider the absolute values of some parameters independently from the values of the others. They present the following justification to this claim [7]: *"For example it is possible that the value of the damage per se (in case of failure) is too high to choose a given decision branch, and this independently either from the probability of the failure (even if it is very low) or from the possible payoff (even if it is very high). In other words, that danger might seem to the agent an intolerable risk"* .

Inspired by Castelfranchi and Falcone, Jøsang and colleagues [11] argued that an explicit distintion between *context-independent trust*, (which they called *reliability trust*), and *context-dependent trust*, (which they called *decision trust*), should be done when using the term trust. They adopt the definition proposed by Gambetta [8] as the definition of *reliability trust*.

Definition 3 (Reliability Trust)
Reliability trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends;

and before introducing the concept of *decision trust*, they propose the following example which can help us to understand the difference between *reliability trust* and *decision trust*:

Consider a person who distrusts an old rope for climbing from the third floor of a house during a fire exercise. Imagine now that the same person is trapped in a real fire in the same house, and that the only escape is to climb from the third floor window with the same old rope. In a real fire, most people would trust the rope. Although the reliability trust in the rope is the same in both situations, the decision trust changes as a function of the utility values associated with the possible courses of action.

Decision trust is then defined as:

Definition 4 (Decision Trust)
Decision trust is the extent to which a given party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible.

The relation between trust and risk in decision making is also considered by Jøsang and Lo Presti in [10]. They propose one of the first models for trust in which a relation between *trust* and *risk* is considered explicitely. Their paper analyses the relationship between the two concepts by first looking at how a decision is made to enter into a transaction based on the risk information.

2.2 . . . to Distrust

While some researchers believe distrust simply means a low level of trust, others believe distrust is a concept entirely separate from trust. Here, we propose a non-exhaustive description of some works which share to some extent our viewpoint on trust and distrust.

Lewicki and colleagues [12] proposed a theoretical framework for understanding simultaneous trust and distrust within relationships. They assert that both trust and distrust involve movements toward certainty: trust concerning expectations of things hoped for and distrust concerning expectations of things feared. Like us, they belong to the thinking school which considers that *trust and distrust are separate but linked dimensions*, but not necessarily the opposite ends of a single continuum. Indeed, the elements which contribute to the growth and decline of trust can be different from those which contribute to the growth and decline of distrust.

Griffiths [9] shown how agents can use trust to manage risk when cooperating. He proposed an approach which (i) uses fuzzy logic to represent trust and distrust; and (ii) allows agents to reason with uncertain and imprecise information regarding other's trustworthiness. He claimed that distrust is not simply a negation of trust, but rather, an explicit belief that an agent will act against the best interest of another. This is in line with Lewicki and colleagues' opinion and also with our opinion. The difference from our opinion is that we consider that distrust should not be always perceived as a reason to necessarily associate malicious intentions to the trustee. In case of interaction with a neutral trustee, that is, when the weight of the trustor's reasons to trust is the same as the weight of the trustor's reasons to distrust, automatically associating malicious (or helpful) purposes to the trustee would not be fair. In that case, decisions should be taken based upon other parameters. For example, an optimistic trustor would underestimate the reasons to distrust: *"an optimist is one who will look for the best in those with whom s(he) interacts"*[13]. A pessimistic trustor instead, would underestimate the reasons to trust.

McKnight and Cheverny [15] argued that trust and distrust are separate constructs that may exist simultaneously. They claimed that *"distrust is not only important because it allows one to avoid negative consequences, but because general distrust of other people and institutions is becoming more prevalent which means that it may, to an extent, be displacing trust as a social mechanism for dealing with risk. Indeed, under certain conditions, distrust may already be more useful or beneficial than trust."* They underline that *"without properly defining trust and distrust, it would be hard to tell which is more important and when."*

3 Towards an Explicit Representation of Distrust

Existing computational models usually deal with trust in a binary way: they assume that a source is either to be trusted or not, and they compute the probability that the source can be trusted. However, sources can not always be divided into trustworthy and untrustworthy in a clear-cut way. Some sources may be trusted to a certain extent. To take this fact into account, we think that trust and distrust should be represented as fuzzy degrees.

3.1 Basic Considerations

Fuzzy sets, introduced by Zadeh [19], are a generalization of classical sets obtained by replacing the characteristic function of a set A with a *membership function* μ_A, which

can take up any value in $[0, 1]$. Let X be the universe of discourse and $x \in X$. The value $\mu_A(x)$ or, more simply, $A(x)$ is the membership degree of element x in A, i.e., the degree to which x belongs in A.

In [1], Atanassov extended the fuzzy set theory by introducing Intuitionistic Fuzzy Set (IFS for short) theory. In fuzzy set theory, it is implicitely assumed that the fact that an element x "belongs" with a degree $\mu_A(x)$ in a fuzzy set A, follows that x should "not belong" to A to the extent $1 - \mu_A(x)$. An intuitionistic fuzzy set F, instead, explicitely assigns to each element x of the considered universe both a degree of membership $\mu_F(x) \in [0, 1]$ and one of non-membership $\nu_F(x) \in [0, 1]$ which are such that:

$$\mu_F(x) + \nu_F(x) \leq 1.$$

Obviously, when $\mu_F(x)+\nu_F(x) = 1$ for all the elements of the universe, the traditional fuzzy set concept is recovered.

Deschrijver and Kerr showed in [6] that IFS theory is formally equivalent to Interval Valued Fuzzy Set (IVFS) theory which is another extention of fuzzy set theory in which the membership degrees are subintervals instead of numbers from $[0, 1]$ [18]. The IFS pair $(\mu_F(x), \nu_F(x))$ corresponds to the IVFS interval $[\mu_F(x), 1 - \nu_F(x)]$, indicating that the degree with which x "belongs in F is ranged from $\mu_F(x)$ to $1 - \nu_F(x)$. They defined the *hesitation degree*, $h \in [0, 1]$, as the length of such an interval. It is given by $h = 1 - \mu_F(x) - \nu_F(x)$. The longer the interval, the more doubt about the actual $\mu_F(x)$ value.

3.2 The Trustworthiness of a Source

The *trustworthiness* of a source (or of another agent) may be defined as, [4]:

Definition [Trustworthiness of a Source]
Let $t \in [0, 1]$ be the degree of trust the agent has in a source, and $d \in [0, 1]$ be its degree of distrust in the same source. The trustworthiness of that source for the agent is represented by pair (t, d), whith $t + d \leq 1$.

Following Deschrijver and Kerr's viewpoint, the trustworthiness (t, d) of a source corresponds to the interval $[t, 1 - d]$, indicating that the trust degree can range from t to $1 - d$. Therefore, the hesitation degree $h = 1 - t - d$ represents the uncertainty, or doubt, about the actual trust value. E.g., if a source has trustworthiness $(0.2, 0)$, this means that the agent trusts the source to degree 0.2, but possibly more, because there is much doubt ($h = 0.8$). More precisely, it means that the agent may trust the source to a degree varying from 0.2 to 1. Instead, if the trustworthiness is $(0.6, 0.4)$, the agent trusts the source to degree 0.6 but not more ($h = 0$).

Thanks to these considerations, we can represent the trustworthiness of a source more faithfully than as it is proposed in existing approaches. For example, we can explicitly represent the following cases of trustworthiness:

$(0, 1)$: the agent has reasons to fully distrust the source, hence it has no hesitation ($h = 0$),
$(0, 0)$: the agent has no information about the source and hence no reason to trust the source, but also no reason to distrust it; therefore, it fully hesitates in trusting it ($h = 1$),

$(1, 0)$: the agent has reasons to fully trust the source, hence it has no hesitation ($h = 0$).

As we can see, by considering both (and not neccessarily related concepts) trust and distrust, it is possible to differentiate between absence of trust caused by presence of distrust (e.g., information provided by a malicious source) versus by lack of knowledge (e.g., as towards an unknown source).

The sources can be classified in:

- *helpful source*: a source for which the reasons to believe in are stronger than the reasons to reject its information;
- *malicious source*: a source for which the reasons to reject its information are stronger than the reasons to believe it;
- *unknown source*: a source which never provided information to the agent before;
- *neutral source*: a source which provided to the agent as much true information as false.

Example Contiunued To sum up, John got information from four sources with different scores. The first source is friendly and competent; therefore, its score is $(1, 0)$. The second is supposedly competent and hopefully independent: therefore, its score might be something like $(\frac{1}{2}, \frac{1}{4})$. The third source is unknown, but has an obvious conflict of interest; therefore John assigns it a score of $(0, 1)$. Finally, the guy on the bus is a complete stranger reporting the opinion of another complete stranger. Therefore, its score cannot be other than $(0, 0)$.

4 Summary and Perspectives for a Normative Multiagent System

Taking distrust explicitly into account can help when making decisions in a situation where the agents are *collaborative*, that is, those which are considered as helpful sources; *wary*, which are suited to contexts where competition is the main theme; and *utility-driven*, for which a gain corresponds to a loss for its counterparts.

It would be interesting to take these considerations into account in the case of a Normative Multiagent System, where the behaviour of an agent depends on its internal components but also on the society it is part of.

Luck and colleagues [16], for example, proposed to analyse the agent's behaviour (reasoning) thanks to a three-dimensional space model, Figure 1, with motivations (axis x), norms (axis y), and trust (axis z). Each vertex in the space represents a kind of society. In particular, an increase in the value of x represents a prevalence of malicious motivations, indicating that agents are more likely to defect if they see more utility in alternative interactions; an increase in the value of y indicates the prevalence of stricter norms and enforcement which can constraint the motivations of agents and prevent them from acting maliciously if they intend to do so; finally, an increase in the value of z indicates an increase in the trust that agents place in other agents and, therefore, an increase in willingness to cooperate with others.

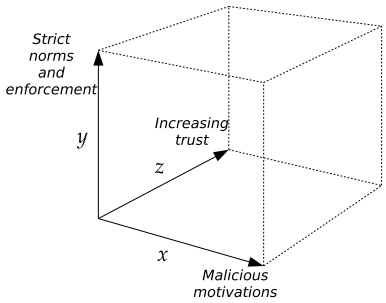

Fig. 1. Three-dimensional space model with norms, motivations and trust.

References

1. K T Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets Syst.*, 20(1):87–96, 1986.
2. C. Castelfranchi and R. Falcone. Trust is Much More Than Subjective Probability: Mental Components and Sources of Trust. In *HICSS'00*. 2000.
3. P. Cofta. Distrust. In *ICEC'06*, pages 250–258. 2006.
4. M. De Cock and P. Pinheiro da Silva. A Many Valued Representation and Propagation of Trust and Distrust. In *WILF'05*, pages 114-120. 2006.
5. R. Demolombe. Reasoning About Trust: A Formal Logical Framework. In *iTrust*, pages 291–303. 2004.
6. G. Deschrijver and E. E. Kerre. On the relationship between some extensions of fuzzy set theory. *Fuzzy Sets Syst.*, 133(2):227–235, 2003.
7. R. Falcone and C. Castelfranchi. Social Trust: A Cognitive Approach. In *Trust and Deception in Virtual Societibes*, pages 55–90. 2001.
8. D. Gambetta. Can we trust trust. In *Trust: Making and Breaking Cooperative Relations*, pages 213–237. 1988.
9. N. Griffiths. A fuzzy approach to reasoning with trust, distrust and insufficient trust. In *CIA*, pages 360–374. 2006.
10. A. Jøsang and S. Lo Presti. Analysing the Relationship Between Risk and Trust. In *Proc. of iTrust'04*, pages 135–145. 2004.
11. A. Jøsang and C. Keser and T. Dimitrakos. Can We Manage Trust. In *Proc. of iTrust'05*, pages 93–107. 2005.
12. R.J. Lewicki and D.J. McAllister, and R.J. Bies. Trust and Distrust: New Relationships and Realities. pages 23(3):438–458. 1998.
13. S. Marsh. Optimism and pessimism in trust. In *Proc. of IBERAMIA'94*, 1994.
14. D. H. McKnight and N. L. Chervany. The Meanings of Trust. University of Minnesota, Management Information Sytems Recherch Center. 1996.
15. D. H. McKnight and N. L. Chervany. Trust and Distrust Definitions: One Bite at a Time. In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies*, pages 27–54. 2001.
16. M. Luck and S. Munroe and F. Lopez y Lopez and R. Ashri. Trust and Norms for Interaction. In *Proc. of the IEEE International Conference on Systems, Man & Cybernetics*, pages 1944–1949. 2004.
17. S. Toivonen and G. Lenzini and I. Uusitalo Context-aware trustworthiness evaluation with indirect knowledge. In *In Proc. of 2nd International Semantic Web Policy Workshop (SWPW'06)*, 2006.

18. I. B. Türksen. Interval valued fuzzy sets based on normal forms. *Fuzzy Sets Syst.*, 20(2):191–210, 1986.
19. L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.