Information sources: Trust and meta-trust dimensions

Cristiano Castelfranchi
ISTC-CNR
Roma, Italy
cristiano.castelfranchi@istc.cnr.it

Rino Falcone
ISTC-CNR
Roma, Italy
rino.falcone@istc.cnr.it

Alessandro Sapienza
ISTC-CNR
Catania, Italy
alessandro.sapienza@istc.cnr.it

Institute for Cognitive Sciences and Technologies, ISTC-CNR

Abstract

We start from the claim that trust in information sources is in fact just a kind of social trust. We are interested in the fact that the relevance and the trustworthiness of the information acquired by an agent X from a source F, strictly depends and derives from the X's trust in F with respect the kind of that information. Even if, of course, we have also to consider the potential (positive and negative) interferences between this information and the previous X's beliefs. In this paper, we analyze the different dimensions of trust in information sources and formalize the degree of subjective certainty or strength of the X's belief P, considering three main factors: the X's trust about P just depending from the X's judgement of the source's competence and reliability; the sources' degree of certainty about P; and the X's degree of trust that P derives from that given source. Finally we present a computational approach based on fuzzy sets and some interesting examples of significant cases.

1 Premise

In our perspective [Castelfranchi and Falcone, 2010; Castelfranchi et al, 2003] trust in information sources is just a kind of social trust, preserving all its prototypical properties and dimensions; just adding new important features and dynamics. In particular, also the trust in information sources [Demolombe, 1999] can just be an evaluation, judgment and feeling, or be a decision to rely on, and act of *believing* in and to the trustee (Y) and rely on it.

Also this trust and the perceived trustworthiness of Y, has *two main dimensions*: the ascribed competence versus the ascribed willingness (intentions, persistence, reliability, honesty, sincerity, etc.).

Also this form of trust is not empty, without a, more or less specified, object/argument: "X trusts Y"! As we have shown, trust is *for/about* something, it has an object: what X expects from Y; Y's service, action, provided good. And it is also *context-dependent*: in a given situation; with internal or external *causal attribution* in case

In: R. Cohen, R. Falcone and T. J. Norman (eds.): Proceedings of the 17th International Workshop on Trust in Agent Societies, Paris, France, 05-MAY-2014, published at http://ceur-ws.org

of success or failure (trust in the agent versus trust in the environment). Also this form of trust is gradable: "X trusts *more or less* Y".

What changes is just the service and good X is expecting from Y, that is *reliable knowledge*. Thus all those dimensions are specified in this direction and acquire special qualities. For example, we can rely on Y without directly knowing it (and without specific recommendation from others) just because Y inherits the trust we have in a given *class* of people or *group* (if Y is assumed to be a good member, a typical instance of it: degree of) [Falcone et al, 2013]. In a practical domain if I trust firefighters or specialized plumbers I can trust Y's *practical* ability as a firefighter, as a professional plumber; analogously, if I trust medical doctors' competence, or press agencies, I will believe the recommendations of this doctor or the news of this press agency. Information is a resource, like others; more or less relevant [Paglieri and Castelfranchi, 2012] and good or bad for our goals. And providing relevant information is a "service" like others.
In particular, in this paper we are interested in the fact that the relevance and the trustworthiness of the information acquired by an agent X from a source F, strictly depends and derives from the X's trust in F with respect the kind of that information.

2 Dimensions of Trust in Information Sources

Given the frame above described, which are the important specific dimensions of trust in information sources (TIS)? Many of these dimensions are quite sophisticated, given the importance of information for human activity and cooperation. We will simplify and put aside several of them.

First of all, we have to trust (more or less) the source (F) as competent and reliable in that domain, in the domain of the specific information content. Am I waiting for some advice on train schedule? On weather forecast? On the program for the examination? On a cooking recipe?
Is this F not only competent but also reliable (in general or specifically towards me)? Is F sincere and honest? Or leaning to lie and deceive? Will F do what has promised to do or "has" to do for his role? And so on.

These competence and reliability evaluations can derive from different reasons, basically:

a) *Direct experience* with F (how F performed in the past interactions) on that specific information content;

b) *Recommendations* (other individuals Z reporting their direct experience and evaluation about F) or *Reputation* (the shared general opinion of others about F) on that specific information content; [Yolum and Singh, 2003; Conte and Paolucci, 2002; Sabater-Mir, 2003; Sabater-Mir and Sierra, 2001; Jiang et al, 2013];

c) *Categorization* of F (it is assumed that a source can be categorized and that it is known this category): on this basis it is also possible to establish the competence/reliability of F on the specific information content [Falcone et al, 2013, Burnett et al, 2010].

The two faces of F's trustworthiness (competence and reliability) are relatively independent[1]; we will treat them as such. Moreover, we will simplify these complex components in just one quantitative fuzzy parameter: F's estimated trustworthiness; by combining competence and reliability.
In particular we define the following fuzzy set: *terrible, poor, mediocre, good, excellent* (see figure 1) and apply it to each of the previous different dimensions (direct experience, recommendations and reputation, categorization).

Second, information sources have and give us a specific information that they know/believe; but believing something is not a yes/no status; we can be more or less convinced and sure (on the basis of our evidences, sources, reasoning). Thus a good source might inform us not only about P, but also about its *degree of certainty about* P, its trust in the truth of P. For example: "It is absolutely sure that P", "Probably P", "It is frequent that P", "It might be that P", and so on.

Of course there are more sophisticated meta-trust dimensions like: how much am I sure, confident, in F's evaluation of the probability of the event or in his subjective certainty? [2] Is F not sincere? Or not so

[1] Actually they are not fully independent. For example, F might be tempted to lie to me if/when is not so competent or providing good products: he has more motives for fudging me.

[2] In a sense it is a *transitivity principle* [Falcone and Castelfranchi, 2012]: X trust Y, and Y trust Z; will X trust Z? Only if X trusts Y "as a good evaluator of Z and of that domain". Analogously here: will X trust Y because Y trusts Y? Only if X trust Y "as a good and reliable evaluator" of it-self.

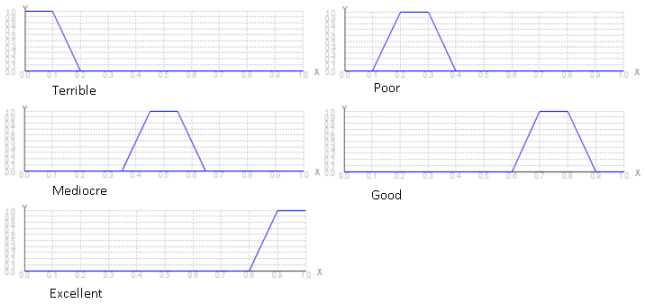

Figure 1: Representation of the five fuzzy sets

self-confident and good evaluator? For example, in drug leaflet they say that a given possible bad side effect is only in 1% of cases. Have I to believe that? Or they are not reliable since they want to sell that drug?
For the moment, we put aside that dimension of how much meta-trust we have in the provided degree of credibility. We will just combine the provided certainty of P with the reliability of F as source. It in fact makes a difference if an excellent or a mediocre F says that the degree of certainty of P is 70% (see §2.2).

Third, especially for information sources it is very relevant the following form of trust: the trust we have that the information under analysis derives from that specific source, how much we are sure about that "transmission"; that is, that the communication has been correct and working (and complete); that there are no interferences and alterations, and I received and understood correctly; that the F is really that F (Identity). Otherwise I cannot apply the first factor: F's credibility.
Let's simplify also these dimensions, and formalize just the *degree of trust that F is F*; that the F of that information (I have to decide whether believe or not) is actually F. In the WEB this is an imperative problem: the problem of the real *identity* of the F, and of the reliability of the signs of that identity, and of the communication.

These dimensions of TIS are quite independent of each other (and we will treat them as such); we have just to combine them and provide the appropriate dynamics. For example, what happen if a given very reliable source F' says that "it is sure that P", but I'm not sure at all that the information really comes from F' and I cannot ascertain that?

2.1 Additional problems and dimensions

We believe in a given datum on the basis of its origin, its source: perception? communication? inference? And so on.

- *The more reliable (trusted) the F the stronger the trust in P, the strength of the Belief that P.*

This is why it is very important to have a "memory" of the sources of our beliefs.

However, there is another fundamental principle of the degree of credibility of a given Belief (its trustworthiness):

- *The many the converging sources, the stronger our belief* (of course, if there are no correlations among the sources).

Thus we have the problem to combine different sources about P, and their subjective degrees of certainty, and their credibility, in order to weigh the credibility of P, and have an incentive due to a large convergence of sources.
There might be different heuristics for dealing with contradictory information and sources. One (*prudent*) agent might adopt as assumption the worst hypothesis, the weaker degree of P; another (*optimistic*) agent, might choose the best, more favorable estimation; another agent might choose the most reliable source. We will formalize only one strategy: the weighing up and combination of the different strengths of the different sources, avoiding however the psychologically incorrect result of probability values, where by combining different probabilities we always decrease the certainty, it never increases. On the contrary - as we said - convergent sources reinforce each other and make us more certain of that datum.

2.2 Feedback on source credibility/TIS

We have to store the sources of our beliefs because, since we believe on the basis of source credibility, we have to be in condition to adjust such credibility, our TIS, on the basis of the result. If I believe that P on the basis of source F1, and later I discover that P is false, that F1 was wrong or deceptive, I have to readjust my trust in F1, in order next time (or with similar sources) to be more prudent. And the same also in case of positive confirmation.[3]
However, remember that it is well known [Urbano et al, 2009] that the negative feedback (invalidation of TIS) is more effective and heavy than the positive one (confirmation). This asymmetry (the collapse of trust in case on negative experience versus the slow acquisition or increasing of trust) is not specific of trust and of TIS; it is -in our view- basically an effect of a general cognitive phenomenon. It is not an accident or weirdness if the disappointment of trust has much stronger (negative) impact than the (positive) impact of confirmation. It is just a sub-case of the general and fundamental asymmetry of negative vs. positive results, and more precisely of "losses" against "winnings": the well-known Prospect theory [Kahneman and Tversky, 1979]. We do not evaluate in a symmetric way and on the basis of an "objective" value/quantity our progresses and acquisitions versus our failures and wastes, relatively to our "status quo". Losses (with the same "objective" value) are perceived and treated as much more severe: the curve of losses is convex and steep while that of winnings is concave. Analogously the urgency and pressure of the "avoidance" goals is greater than the impulse/strength of the achievement goals [Higgins, 1997]. All this applies also to the slow increasing of trust and its fast decreasing; and to the subjective impact of trust disappointment (betrayal!) vs. trust confirmation. That's why usually we are prudent in deciding to trust somebody; in order do not expose us to disappointment and betrayals, and harms. However, also this is not always true; we have quite naive forms of trust just based on gregariousness and imitation, on sympathy and feelings, on the diffuse trust in that environment and group, etc. This also plays a crucial role in social networks on the web, in web recommendations, etc.

Moreover, in our theory [Falcone and Castelfranchi, 2004] not always and automatically a bad result (or a good result) entails the revision of TIS. It depends on the "causal attribution": it has been a fault/defect of F or an interference on the environment? The result might be bad although F's performance was his best. Let us put aside here the feedback effect and revision on TIS.

2.3 Where does TIS comes from?

Which are the sources, bases, of our trust in our information sources, what determine our trust in their message/content? Let's add something more specific on what we have seen in 1.1, also because in the literature "direct experience" and reputation-recommendation play an absolutely dominant role:

A) Our previous direct experience with F, or better our "memory" about, and the adjustment that we have made of our evaluation of F in several interaction, and possible successes or failure relying on its information.

B) The others' evaluation/trust; either inferred from their behavior or attitude by "transitivity": "If Y trusts Z, me too can trust Z", or due to explicit recommendation from Y (and others) about Z; or due to Z's "reputation" in that social environment: circulating, emergent opinion.

C) Or by inference and reasoning:

   i) by inheritance from classes or groups were Z id belonging (as a good "exemplar");

   ii) by analogy: Z is (as for that) like Y, Y is good for, then Z too is good for;

   iii) by analogy on the task: Z is good/reliable for P he should be good also for P', since P and P' are very similar. (In any case: how much do I trust my reasoning ability?).

2.4 Plausibility: the integration with previous knowledge

To believe something means not just to put it in a file in my mind; it means to *"integrate" it with my previous knowledge*. Knowledge must be at least non-contradictory, and possibly supported, justified: this explains that, and it is explained, supported, by these other facts/arguments. If there is *contradiction* I cannot believe P;

[3] We can even memorize something that we reject. We do not believe to it but not necessarily we delete/forget it, and its source. This is for the same function: in case that information would result correct I have to revise my lack of trust in that source.

either I have to reject P or I have to revise my previous beliefs in order to coherently introduce P. It depends on the strength of the new information (its credibility, due to its sources) and on the number and strength of the internal opposition: the value of the contradictory previous beliefs, and the extension and cost of the required revision. That is: *it is not enough that the candidate belief that P be well supported and highly credible*; is there an epistemic conflict? Is it "implausible" to me? Are there antagonistic beliefs? And which is their strength? The winner of the conflict will be the stronger "group" of beliefs. Even the information of a very credible source (like our own eyes) can be rejected!

2.5 Trusting as Risking

In which sense and on what ground we make "reliance" on a belief? We decide and pursue a given goal precisely on the basis of what we believe, and of the of our certainty on current circumstances and predictions. We reasonably invest and persist in a given activity proportionally to our degree of certainty, of confidence: the more we do believe in it the more we bet on and take risks (we sacrifice alternatives and resources, and expose ourselves to possible failures, wastes, and regrets). In other words, our trust in a given information or fact exposes us to "risks". Trust always means to take same risk. As for TIS our claim is that:

- *the higher the perceived risk* (estimated probability and gravity of the failure plus probability and gravity of the possible harms) *the higher the threshold for accepting a given belief and for deciding (trust as decision) to act on such a basis.*

In this case, we would search or wait for additional data or sources. We will not formalize in this paper, this crucial aspect.

2.6 Layered trust

Notice the recursive and inheritance structure of trust and the various meta-levels: I rely and risk on a given assumption on the basis of how much I believe in it, but I believe in it (trust in Belief) on the basis of the number, convergence, and credibility of the information source (TIS) and on the trust they have in what they "say" (and on the trust I have in the trust they have on what they say: meta-meta-trust). And I trust a given source on the basis of information about it: from my memory and experience, from inference, from others (transitivity, reputation, or recommendation), then I have to trust these new information sources: the information about my information sources. And so on [Castelfranchi and Falcone, 2010].

3 Formalizing and computing the degree of certainty as trust in the belief

As we have said, there is a confidence, a trust in the beliefs we have and on which we rely. Suppose X is a cognitive agent, an agent who has beliefs and goals. Given Bel_X, the set of the X's beliefs, then P is a belief of X if:

$$P \in Bel_X \tag{1}$$

The degree of subjective certainty or strength of the X's belief P corresponds with the X's trust about P, and call it:

$$Trust_X(P) \tag{2}$$

3.1 Its origin/ground

In the case in which we are considering just one information, P, deriving from just one source F, $Trust_X(P)$ depends from:

i) the X's trust towards F as source of the information P (that could mean with respect the class of information to which P belongs):
$$Trust_X(F, P) \tag{3}$$
and

ii) the relationships between P and the other X's beliefs (see §1.5).

With respect to the case (ii), if we term Q a X's belief $Q \in Bel_X$ and $Trust_X(P|Q)$ the X's Trust in P given the X's belief Q; we can say that:

- if $Trust_X(P|Q) > Trust_X(P)$ (4)
  Q positively interferes with P (Q supports P);

- if $Trust_X(P|Q) < Trust_X(P)$ (5)
  Q negatively interferes with P (Q is in contradiction with P).

So we can say that given a X's belief Q, it can have positive, negative or neutral interference with the new information P that X is acquiring. The value of this interference (in the positive or negative cases) will be called $\Delta INT_{X,P}(Q)$:
$\Delta INT_{X,P}(Q) = |Trust_X(P|Q) - Trust_X(P)|$ (6)

It is a function of two main factors, the X's trust in Q and the strength of the Q's interference with P:
$\Delta INT_{X,P}(Q) = f_1(Trust_X(Q), DInt(Q,P))$ (7)

It is important to underline that we are considering the composition among different information (P, Qi) just from a very abstract and not analytical point of view. In this work we do not cope with the relevant problems studied in this domain by the research on argumentation [Walton, 1996; Walton et al; 2008].

So, supposing there is just one X's belief (Q) interfering with P, we have that:
$Trust_X(P) = f_2(Trust_X(F,P), \Delta INT_{X,P}(Q))$ (8)
in words: the X's trust in the information P is a function of both the X's trust in the source F about P and of the interference factor between P and Q (as showed by $\Delta INT_{X,P}(Q)$). In the case of more than one X's beliefs interfering with P, say Q_i (with i= 1, ..., n), we have to compose the n interfering factors ($\Delta INT_{X,P}(Q_i)$) in just one resulting factor.

Applying now the conceptual modeling described in the §1.1 we have that $Trust_X(F,P)$ can be articulated in:

1. the X's trust about P just depending from the X's judgement of the F's competence and reliability as derived from the composition of the three factors (direct experience, recommendation/reputation, and categorization), in practice the F's credibility about P on view of X:
   $Trust^1_X(F,P)$ (9)

2. the F's degree of certainty about P: information sources give not only the information but also their certainty about this information; given that we are interested to this certainty, but we have to consider that through X's point of view, we introduce
   $Trust_X(Trust_F(P))$ (10)
   in particular, we consider that X completely trusts F, so that $Trust_X(Trust_F(P)) = Trust_F(P)$

3. the X's degree of trust that P derives from F: the trust we have that the information under analysis derives from that specific source:
   $Trust_X(Source(F,P))$ (11)

Resuming:
$Trust_X(F,P) = f_3(Trust^1_X(F,P), Trust_X(Trust_F(P)), Trust_X(Source(F,P)))$ (12)
Here we could introduce a threshold for each of these 3 dimensions, allowing to reduce risk factors.

3.2 A modality of computation

3.2.1 $Trust^1_X(F,P)$

As specified in §1.3 the value of $Trust^1_X(F,P)$ is a function of:

1. Past interactions;

2. The category of membership;

3. Reputation.

As previously said, each of these values is represented by a fuzzy set: terrible, poor, mediocre, good, excellent. We then compose them into a single fuzzy set, considering a weight for each of these three parameters. Those weights are defined in range [0;10], with 0 meaning that the element has no importance in the evaluation and 10 meaning that it has the maximal importance. It is worth noting that the weight of experience has to be referred to a twofold meaning: it must take into account the numerosity of experiences (with their positive and negative values), but also the intrinsic value of experience for that subject.
However, the fuzzy set in and by itself is not very useful: what interests us in the end is to have a plausibility range, which is representative of the expected value of $Trust^1_X(F,P)$. To get that, it is therefore necessary to apply a defuzzyfication method. Among the various possibilities (mean of maxima, mean of centers) we have chosen to use the centroid method, as we believed it can provide a good representation of the fuzzy set. The centroid method exploits the following formula:

$$k = \frac{\int_0^1 x f(x)\,\mathrm{d}x}{\int_0^1 f(x)\,\mathrm{d}x}$$

were $f(x)$ is the fuzzy set function. The value k, obtained in output, is equal to the abscissa of the gravity center of the fuzzy set. This value is also associated with the variance, obtained by the formula:

$$\sigma^2 = \frac{\int_0^1 (x-k) f(x)\,\mathrm{d}x}{\int_0^1 f(x)\,\mathrm{d}x}$$

With these two values, we determine $Trust^1_X(F,P)$ as the interval $[k-\sigma; k+\sigma]$.

3.2.2 $Trust_X(F,P)$

Once we get $Trust^1_X(F,P)$, we can determine the value of $Trust_X(F,P)$. In particular, we determine a trust value followed by an interval, namely the uncertainty on $Trust_X(F,P)$. For uncertainty calculation we use the formula:
$Uncertainty = 1 - (1 - \Delta x) * Trust_X(Trust_F(P)) * Trust_X(Source(F,P))$
$\Delta x = Max(Trust^1_X(F,P)) - Min(Trust^1_X(F,P))$
In other words, the uncertainty depended on the uncertainty interval of $Trust^1_X(F,P)$, properly modulated by $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$. This formula implies that uncertainty:

- Increase/decrease linearly when Δx increase/decrease;
- Increase/decrease linearly when $Trust_X(Trust_F(P))$ decrease/increase;
- Increase/decrease linearly when $Trust_X(Source(F,P))$ decrease/increase;

The inverse behavior of $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ is perfectly explained by the fact that when X is not so sure that P derives from F or F's degree of certainty about P is low, global uncertainty should increase. The maximum uncertainty value is 1 (+-50%) meaning that X is absolutely not sure about its evaluation. On the contrary, the minimum value of uncertainty is 0, meaning that X is absolutely sure about its evaluation.

In a way similar to uncertainty, we used the following formula to compute a value of $Trust_X(F,P)$:
$Trust_X(F,P) = 1/2 + (Trust^1_X(F,P) - 1/2) * Trust_X(Trust_F(P)) * Trust_X(Source(F,P))$
This formula has a particular trend, different from that of uncertainty. Here in fact the point of convergence is 1/2, value that does not give any information about how much X can trust F about P. Notice that:

- If $Trust^1_X(F,P)$ is less than 1/2, as $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ increase the value of trust will decrease going to the value of $Trust^1_X(F,P)$; as $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ decrease the value of trust will increase going to 1/2;
- If $Trust^1_X(F,P)$ is more than 1/2, as $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ increase the value of trust will increase going to the value of $Trust^1_X(F,P)$; as $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ decrease the value of trust will decrease going to 1/2.

3.2.3 Computing a final trust value

Here we use an aggregation formula based on the sinusoidal function [Urbano et al, 2009]:
$\gamma = \delta * sin(\alpha) + \delta$

where:
$\delta = 1/2$;
$\omega_0 = \pi$;
$\omega = \omega_0/8$;
$\alpha_0 = 3/2\pi$;
$\alpha = \alpha_0 + \omega_0 * Trust_X(F,P) + PP * \omega * (\Sigma\Delta Supp\_INT_{X,P}(Q)) + PN * \omega * (\Sigma\Delta Opp\_INT_{X,P}(Q))$
ω is the aggregation step. The higher it is the faster a high number of supporting/opposing beliefs let the trust value converge to 1/0 (it depends from the division factor: we chose 8). PP and PN are respectively the weights of positive and negative beliefs. Of course, they are function of the subjective value of the goal and the risk that the trustor is willing to address. However, in this work we assume them to be constant.
PP = 1
PN = 1
Let us specify a methodological aspect. The choice of the parameters of the previous formulas should emerge from the many features of the domain (types of agents, information, sources, and so on) and from experiments simulating them. In fact we are introducing these formulas just considering a top-down model, so our choices of these parameters are quite arbitrary.

How previously stated, interference of a single belief is determined by X's trust in Q and the strength of the Q's interference with P. In particular $\Delta INT_{X,P}(Q) = Trust_X(Q) * DInt(Q,P)$

$DInt(Q,P)$ belongs to the range [-1,1], with a value of -1 meaning totally opposing and 1 totally supporting.

Moreover every interference (actually its absolute value), before being taken into account in the final calculation, is compared to a threshold: if the value is less than the threshold it is discarded.

3.3 Some interesting examples

Let us now present some examples of interest.

EXAMPLE 1
In the first example we show hot to compute a value of $Trust^1_X(F,P)$.
In this particular situation, X gives really more importance to past experience (maximum weight value) than to category (medium weight value). Here reputation is not so important.

Table 1: INPUT DATA

Reputation	good
Reputation weight	2
Category	mediocre
Category weight	5
Past Experience	excellent
Past Experience weight	10

Table 2: OUTPUT DATA

Mean	0.7511
Variance	0.04119

The result is that we have a quite high value of $Trust^1_X(F,P)$, with low variance.

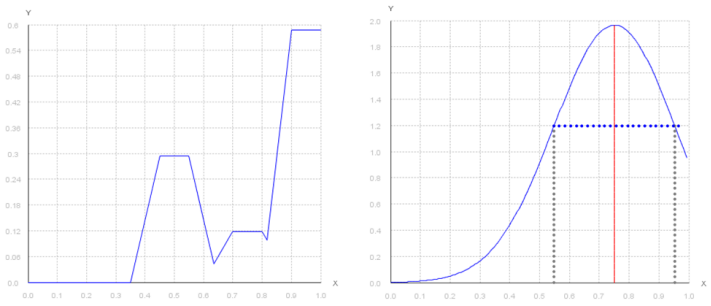

Figure 2: Global Fuzzy set and Gaussian Trend in the example

EXAMPLE 2
In a second situation, X has not past experience to use in his evaluation, but it can still assess a value of $Trust^1_X(F, P)$, basing its evaluation on reputation and category. Supposing that it is in an environment in which there is a high reputation, it decides to give more importance to this factor.

Table 3: INPUT DATA

Reputation	excellent
Reputation weight	10
Category	mediocre
Category weight	2
Past Experience	none
Past Experience weight	–

Table 4: OUTPUT DATA

Mean	0.83333
Variance	0.0323

Here are the results:

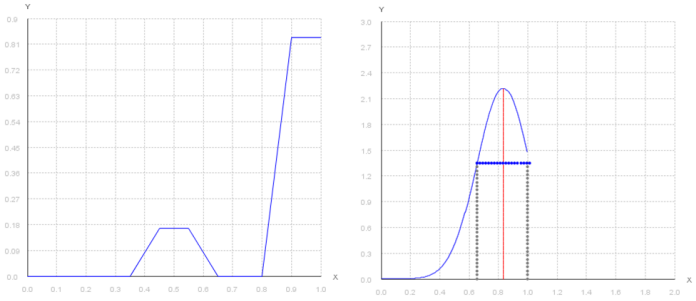

Figure 3: Global Fuzzy set and Gaussian Trend in the example

EXAMPLE 2.1
Now, fixing the value of $Trust^1_X(F,P)$ from the last example, let's see how $Trust_X(Trust_F(P))$ and $Trust_X(Source(F,P))$ influence the value of $Trust_X(F,P)$.

Table 5: INPUT 1

$Trust_X(Trust_F(P))$	1
$Trust_X(Source(F,P))$	1

Table 6: OUTPUT 1

$Trust_X(F,P)$	0.83333
Uncertainty	+-17.32%

In this case, the source F is sure about the belief P, then $Trust_X(Trust_F(P)) = 1$. Moreover X is sure that the source of P is exactly F, then $Trust_X(Source(F,P)) = 1$. This implies that $Trust_X(F,P)$ is exactly equal to $Trust^1_X(F,P)$.

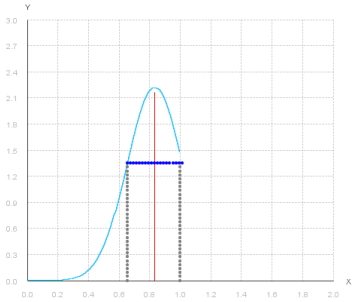

Figure 4: Gaussian Trend

EXAMPLE 2.2
Here X evaluates an uncertainty of F about the belief P.

Table 7: INPUT 2

$Trust_X(Trust_F(P))$	0.5
$Trust_X(Source(F,P))$	1

Table 8: OUTPUT 2

$Trust_X(F,P)$	0.66666
Uncertainty	+-33.66%

We can see a tendency of $Trust_X(F,P)$ to decrease (such decrease has as a lower limit the maximal trust ambiguity, that is 0.5) and uncertainty increases towards its maximal value, that is +-0.5.

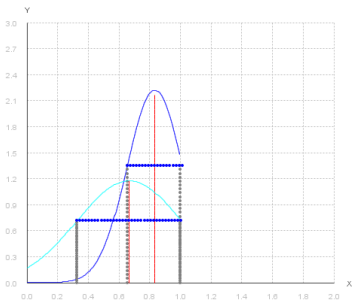

Figure 5: Gaussian Trend

EXAMPLE 2.3

Table 9: INPUT 3

$Trust_X(Trust_F(P))$	0.5
$Trust_X(Source(F,P))$	0

Table 10: OUTPUT 3

$Trust_X(F,P)$	0.5
Uncertainty	+-50.0%

Here X is in the worst situation, it cannot assume anything. This makes perfect logical sense because it is sure F is not the source of P.

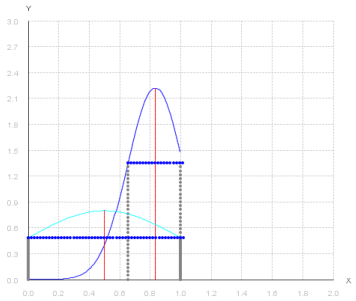

Figure 6: Gaussian Trend

4 Conclusions

We would like underline that the main differences with our previous works on this topic (Castelfranchi et al, 2003] have to be referred to a new revised conceptualization and formalization of the trust model in information sources and in particular to the set of simulations designed for evaluating how the different components of trust (in the information source) interfere among themselves and how they contribute to the final value of trust in that information (P).

Let us conclude with the "Liar Paradox": A specific distrust in F's sincerity and honesty, that is the strong belief that F is lying while asserting that P, has a paradoxical consequence: *I can reasonably come to believe the opposite*. If F is lying and thus P is false, it is true (I have to believe) that Not (P).
Not necessarily this means that I know how the world is. If there is a two-value world/case (P or Q), given that P is false I'm sure that Q (I can trust the belief that Q given my distrust in F!). But, if the case/world has

multiple alternative (P or Q or W or Z), I can just believe (with certainty) that given Not (P) it is either Q or W or Z.
Two burglars are trying to break a shop open when the police shows up. One burglar manages to fly away, but the other is caught by the policemen, who ask him which way did his accomplice go. "That way!", answers the burglar, while pointing to the right. And the policemen run in the opposite direction!
Notice that when my distrust is about the *competence* dimension, that is I'm sure that F is not expert and informed at all about P, F's assertion that "surely P" doesn't give me the certainty that Not (P), but just leave me fully *uncertain*: I don't know, I cannot use F as source for P.
This different effect (especially for TIS) between the *competence dimension* of trustworthiness and the *honesty/sincerity* (reliability) dimension is quite interesting and can help to clarify the necessity of a rich analysis of the trust in information sources.

5 Acknowledgments

This work is partially supported by the Project PRISMA (PiattafoRme cloud Interoperabili per SMArtgovernment; Cod. PON04a2_A) funded by the Italian Program for Research and Innovation (Programma Operativo Nazionale Ricerca e Competitività 2007-2013).

References

[Castelfranchi and Falcone, 2010] Castelfranchi C., Falcone R., Trust Theory: A Socio-Cognitive and Computational Model, John Wiley and Sons, April 2010.

[Castelfranchi et al, 2003] Castelfranchi, C., Falcone R., Pezzulo, (2003) Trust in Information Sources as a Source for Trust: A Fuzzy Approach, Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-03) Melburne (Australia), 14-18 July, ACM Press, pp.89-96.

[Falcone et al, 2013] Falcone R., Piunti, M., Venanzi, M., Castelfranchi C., (2013), From Manifesta to Krypta: The Relevance of Categories for Trusting Others, in R. Falcone and M. Singh (Eds.) Trust in Multiagent Systems, ACM Transaction on Intelligent Systems and Technology, Volume 4 Issue 2, March 2013

[Paglieri and Castelfranchi, 2012] Paglieri F., & Castelfranchi C. (2012). Trust in relevance. In: S. Ossowski, F. Toni, G. Vouros (Eds.), Proceedings of the First International Conference on Agreement Technologies (AT 2012). CEUR Workshop Proceedings, vol. 918: CEUR-WS.org, pp. 332 - 346.

[Yolum and Singh, 2003] Yolum, P. and Singh, M. P. 2003. Emergent properties of referral systems. In Proceedings of the 2nd International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS'03).

[Conte and Paolucci, 2002] Conte R., and Paolucci M., 2002, Reputation in artificial societies. Social beliefs for social order. Boston: Kluwer Academic Publishers.

[Sabater-Mir, 2003] Sabater-Mir, J. 2003. Trust and reputation for agent societies. Ph.D. thesis, Universitat Autonoma de Barcelona.

[Burnett et al, 2010] Burnett, C., Norman, T., and Sycara, K. 2010. Bootstrapping trust evaluations through stereotypes. In Pro- ceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10). 241248.

[Falcone and Castelfranchi, 2012] Falcone R., Castelfranchi C., Trust and Transitivity: How trust-transfer works, 10th International Conference on Practical Applications of Agents and Multi-Agent Systems, University of Salamanca (Spain)28-30th March, 2012.

[Urbano et al, 2009] Joana Urbano, Ana Paula Rocha, and Eugnio Oliveira, Computing Confidence Values: Does Trust Dynamics Matter? In L. Sabra Lopes et al. (Eds.): EPIA 2009, LNAI 5816, pp. 520-531, 2009, Springer.

[Kahneman and Tversky, 1979] Kahneman, Daniel, and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk". Econometrica. XLVII (1979): 263-291.

[Higgins, 1997] Higgins, E. T. (1997). Beyond pleasure and pain. American Psychologist, 52, 1280-1300.

[Falcone and Castelfranchi, 2004] Falcone R., Castelfranchi, C. (2004), Trust Dynamics: How Trust is influenced by direct experiences and by Trust itself; Proceedings of the 3rd International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-04), New York, 19-23 July 2004, ACM-ISBN 1-58113-864-4, pages 740-747.

[Sabater-Mir and Sierra, 2001] Sabater-Mir J., Sierra C., (2001), Regret: a reputationmodel for gregarious societies. In 4th Workshop on Deception and Fraud in Agent Societies (pp. 61-70). Montreal, Canada.

[Demolombe, 1999] Demolombe R., (1999), To trust information sources: A proposal for a modal logic framework. In Castelfranchi C., Tan Y.H. (Eds), Trust and Deception in Virtual Societies. Kluwer, Dordrecht.

[Jiang et al, 2013] S. Jiang, J. Zhang, and Y.S. Ong. An evolutionary model for constructing robust trust networks. In Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2013.

[Walton et al, 2008] D. Walton, C. Reed and F. Macagno, Argumentation Schemes, Cambridge, Cambridge University Press, 2008.

[Walton, 1996] D. Walton, Argumentation Schemes for Presumptive Reasoning , Mahwah, N.J., Lawrence Erlbaum Associates, 1996.