

# How to manage the information sources' trustworthiness in a scenario of hydrogeological risks

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

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

## Abstract

In this work we present a study about cognitive agents that have to learn how their different information sources can be more or less trustworthy in different situations and with respect to different hydrogeological phenomena. We introduced an ad-hoc Bayesian trust model that we created and used in the simulations. We also describe the realized platform that can be manipulated in order to shape many possible scenarios. The simulations are populated by a number of agents that have three information sources about forecasts of different hydrogeological phenomena. These sources are: a) their own evaluation/forecast about the hydrogeological event; b) the information about the event communicated by an authority; c) the behavior of other agents as evidence for evaluating the dangerous level of the coming hydrogeological event. These weather forecasts are essential for the agents in order to deal with different and more or less dangerous meteorological events requiring adequate behaviors. We consider in particular in this paper some specific situations in which the authority can be more or less trustworthy and more or less able to deliver its own forecasts to the agents. The simulations will show how, on the basis of a training phase in these different situations, the agents will be able to make a rational use of their different information sources.

## 1 Introduction

One of the main problems we have for understanding and foreseeing any domain and world, is not just to access to the different information sources about that domain, but also to be able to evaluate the trustworthiness of those sources.

In particular, the same source not necessarily has the same degree of trustworthiness in any situation and context, but could change its own reliability on the basis of different external or internal factors. For example, in the domain of the weather forecasts, we know that the mathematical models used for this task are quite reliable when referring to temporally close events (next 6-24 hours), while are quite approximate if referring to

---

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

In: J. Zhang, R. Cohen, and M. Sensoy (eds.): Proceedings of the 18th International Workshop on Trust in Agent Societies, Singapore, 09-MAY-2016, published at <http://ceur-ws.org>

long-term events (1-4 weeks). And we also know that these forecasts can change their reliability on the basis of the kind of phenomenon they are evaluating.

So can be very relevant to have different information sources and also to know how trustworthy they are in different contexts and situations.

On the other hand, trying to put together information coming from different information sources can be an uneasy task. It is necessary to have strategies to do it, especially in presence of critical situation, when there are temporal limits to get decision and a wrong choice can lead to an economical loss or even to risk life.

As said, the necessity of integrating sources on different scopes can be very useful in order to make a well-informed decision. In case of the weather forecast we can consider different sources: official bulletin of authorities, the observation of other agents behavior and of their decisions during the meteorological event, the direct evaluation and competence of the same agents as the basis for their own decisions.

Some of these sources are not correlated among them: a forecast is referred to mathematical model of the weather linked to its previous data, while a direct evaluation can be based on a current human perception of the phenomenon (with its potential psychological and perceptive bias). Then, integrating these sources becomes essential and at the same time it is necessary to identify and take into account their trustworthiness.

In our view [Castelfranchi and Falcone, 2010] trusting an information source (S) means to use a cognitive model based on the dimensions of competence and reliability/motivation of the source. These competence and reliability evaluations can derive from different reasons, basically:

- Our previous *direct experience* with S on that specific kind of information content.
- *Recommendations* (other individuals Z reporting their direct experience and evaluation about S) or *Reputation* (the shared general opinion of others about S) on that specific information content [Conte and Paolucci, 2002; Jiang et al, 2013; Sabater-Mir, 2003; Sabater-Mir and Sierra, 2001; Yolum and Singh, 2003].
- *Categorization* of S (it is assumed that a source can be categorized and that it is known this category), exploiting inference and reasoning (analogy, inheritance, etc.): on this basis it is possible to establish the competence/reliability of S on that specific information content [Burnett et al, 2010; Burnett et al, 2013; Falcone and Castelfranchi, 2008; Falcone et al, 2013].

However in this paper, for sake of simplicity, we use just the first kind of the three reasons above described: the direct experience with each source.

Our agents do manipulate their values of trust (we consider the feedback effects and the trust dynamics). In practice each agent evaluates if an information source was corrected with its own prediction and, on the basis of this evaluation, decides to increase its trustworthiness or decrease it.

We start from agents equally trusting the three different sources (neutral agents) and then apply to them a training period experimenting different weather scenarios. We do this with the presence of four kinds of authorities: reliable and strongly communicative, reliable and weakly communicative, not reliable and strongly communicative, and not reliable and weakly communicative.

Our investigation is about: are the agents able to learn the more trustworthy sources? Are they able to intelligently integrate these sources? Are the agents performances coherent with the trustworthiness of the sources they are following? Are we able to extract useful information from these simulations for situations of real cases? In the paper we show how, with a certain limit of approximation, we can give some useful and interesting indications.

## 2 The trust model

Given the complexity of simulations, we chose to use a relatively simple trust model, unifying many parameters in just one.

Trust decision in presence of uncertainty can be handle using uncertainty theory [Liu, 2014] or probability theory. We decided to use the second approach, as in this platform agents know a priori all the possible events that can happen and they are able to estimate how much it is plausible that they occur. In particular we exploit Bayesian theory, one of the most used approach in trust evaluation [Melaye and Demazeau, 2005; Quercia et al, 2006; Wang and Vassileva; 2003].

In this model each information source S is represented by a trust degree called *TrustOnSource*, with  $0 \leq \textit{TrustOnSource} \leq 1$ , plus a bayesian probability distribution PDF<sup>1</sup> (Probability Distribution Function) that represents the information reported by S.

The trust model takes into account the possibility of many events: it just split the domain in the corresponding number of interval. In this work we use three different events (described below), then the PDF will be divided into three parts.

The *TrustOnSource* parameter is used to smooth the information referred by S. This is the formula used for transforming the reported PDF:

$$\textit{NewValue} = 1 + (\textit{Value} - 1) * \textit{TrustOnSource}$$

The output of this step is called Smoothed PDF (SPDF). We will have that:

- The greater *TrustOnSource* is, the more similar the SPDF will be to the PDF; in particular if  $\textit{TrustOnSource} = 1 \Rightarrow \textit{SPDF} = \textit{PDF}$ ;
- The lesser it is, the more the SPDF will be flatten; in particular if  $\textit{TrustOnSource} = 0 \Rightarrow \textit{SPDF}$  is an uniform distribution with value 1.

The idea is that we trust on what S says proportionally to how much we trust S. In words, the more we trust S, the more we tend to take into consideration what it says; the less we trust S, the more we tend to ignore its informative contribution.

We define GPDF (Global PDF) the evidence that an agent owns concerning a belief P. Once estimated the SPDFs for each information source, there will be a process of aggregation between the GPDF and the SPDFs. Each source actually represents a new evidence E about a belief P. Then to the purpose of the aggregation process it is possible to use the classical Bayesian logic, recursively on each source:

$$f(PE) = (f(EP) * f(P))/f(E)$$

where:  $f(P|E) = \textit{GPDF}$  (the new one)

$f(E|P) = \textit{SPDF}$ ;

$f(P) = \textit{GPDF}$  (the old one)

In this case  $f(E)$  is a normalization factor, given by the formula:

$$f(E) = \int f(EP) * f(P)dP$$

In words the new GPDF, that is the global evidence that an agent has about P, is computed as the product of the old GPDF and the SPDF, that is the new contribute reported by S. As we need to ensure that GPDF is still a probability distribution function, it is necessary to scale it down<sup>2</sup>. This is ensured by the normalization factor  $f(E)$ .

---

<sup>1</sup>It is modeled as a distribution continuous in each interval

<sup>2</sup>To be a PDF, it is necessary that the area subtended by it is equal to 1.

## 2.1 Feedback on trust

We want to let agents adapt to the context in which they move. This means that, starting from a neutral trust level (that does not imply trust or distrust) agents will try to understand how much to rely on each single information source.

To do that, they need a way to perform feedback on trust. We propose to use weighted mean. Given the two parameters  $\alpha$  and  $\beta^3$ , the new trust value is computed as:

$$newTrustDegree = \alpha * oldTrustDegree + \beta * performanceEvaluation$$

$$\alpha + \beta = 1$$

Where oldTrustDegree is the previous trust degree and the performanceEvaluation is the objective evaluation of the source performance. This last value is obtained comparing what the source said with what actually happened. Considering the PDF reported by the source, and remembering that it is split into three parts, we will have that:

1. The estimated probability of the event that actually occurred is completely taken into account;
2. The estimated probability of the event immediately near to the actual one is taken into account for just 1/3. We in fact suppose that even if the evaluation is not right, it is not even completely wrong.
3. The rest of the PDF is not considered.

Lets suppose that there has been a critical event. A first source reported a 100% probability of critical event; a second one a 50% probability of critical event and a 50% of medium event; finally a third one asserts 100% of light event

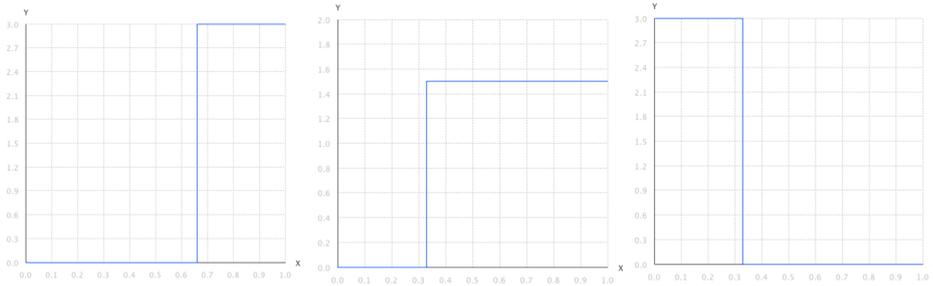


Figure 1: (a) A source reporting a 100% probability of critical event. (b) A source reporting a 50% probability of critical event and 50% probability of medium event. (c) A source reporting a 100% probability of light event

The performance evaluation for them will be:

1. Source1: 100%
2. Source2:  $50\% + (50/3)\% = 66.67\%$
3. Source3: 0%

## 3 The platform

Exploiting NetLogo [Wilensky, 1999], we created a very flexible platform, where a lot of parameters are taken into account to model a variety of situations.

<sup>3</sup>Of course changing the values of  $\alpha$  and  $\beta$  will have an impact on the trust evaluations. With high values of  $\alpha/\beta$ , agents will need more time to get a precise evaluation, but a low value (below 1) will lead to an unstable evaluation, as it would depend too much on the last performance. We do not investigate these two parameters in this work, using respectively the values 0.9 and 0.1. In order to have good evaluations, we let agents make a lot of experience with their information sources.

### 3.1 The context

The basic idea is that, given a population distributed over a wide area, some weather phenomena happen in the world with a variable level of criticality. These weather phenomena happen in the world in a temporal window of 16 ticks.

The world is populated by a number of cognitive agents (citizens) that react to these situations, deciding how to behave, on the basis of the information sources they have and of the trustworthiness they attribute to these different sources: they can escape, take measures or evaluate absence of danger.

In addition to citizens, there is another agent called authority. Its aim is to inform promptly citizens about the weather phenomena. The authority will be characterized by an uncertainty, expressed in terms of standard deviation, and by a communicativeness value, that represents the probability that it will be able to inform each single citizen.

### 3.2 Information Sources

To make a decision, each agent can consult a set of information sources, reporting to it some evidence about the incoming meteorological phenomena. We considered the presence of three kinds of information sources (whether active or passive) for agents:

1. Their *personal judgment*, based on the direct observation of the phenomena. Although this is a direct and always true (at least in that moment) source, it has the drawback that waiting to see what happens could lead into a situation in which it is no more possible to react in the best way (for example there is no more time to escape if one realizes too late the worsening weather).
2. *Notification from authority*: the authority distributes into the world weather forecast with associated different alarm signals, trying to prepare citizens to what is going to happen. It is not sure that the authority will be able to inform everyone
3. *Others behavior*: agents are in some way influenced by community logics, tending to partially or totally emulate their neighbors behavior.

The notification from the authority is provided as a clear signal: all the probability is focused on a single event. Conversely, the personal judgment can be distributed on two or three events with different probabilities. This can also be true for others behavior estimation as the probability of each event is directly proportional to the number of neighbors making each kind of decision. If no decision is available, the PDF is a uniform distribution with value 1.

### 3.3 Agents' Description

At the beginning of the simulation, the world is populated by a number of agents. These agents have the same neutral trust value 0.5 for all their information sources. This value represents a situation in which agents are not sure if to trust or not a given source, as a value of 1 represents complete trust and 0 stands for complete distrust.

Agents are also characterized by a decision deadline, expressed in ticks, that determines the moment in which agents will make a decision.

The unique difference between them relies on how much they are able to see and to read the phenomena. In fact, in the real world not all the agents have the same abilities. In order to shape this, we divided equally agents into three set:

1. Good evaluators: agents that are able to see 15 ticks of event. They will be quit always able to detect correctly the event, then we expect them to rely mainly on their own opinion.
2. Medium evaluators: agents that are able to see 14 ticks of event. They can detect the event, but not good as the previous category.
3. Bad evaluators: agents that are able to see 13 ticks of event. Quite often, they will detect two possible events, but they will need another source to decide between them.

### 3.4 World Description

The world is made by 32x32 patches, which wraps both horizontally and vertically. It is geographically divided in 4 quadrants of equal dimension, where agents are distributed in a random way.

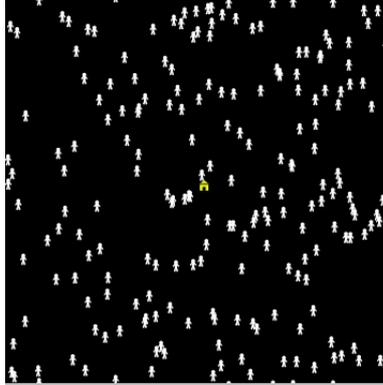


Figure 2: A world example. There are 200 agents plus the authority, represented by the yellow house in the center.

The quadrants differ in the possible weather phenomena that happen, modeled through the presence of clouds. The events are modeled so that agents can't be completely sure of what is going to happen:

1. *Critical event*: a tremendous event due to a very high level of rain, with possible risks for the agents sake; it is presented through a 16 ticks sequence of 3 clouds;
2. *Medium event*: it can cause possible damage to house or streets, but there is not health hazard; it is composed by a 16 ticks sequence of 2 or 3 clouds (or by a sequence of 13 ticks at the beginning followed by at least a couple of (2,3) in any sequence). To let this event be similar to the critical one, the 50% of the times we force the firsts 13 ticks to be equal to 3.
3. *Light event*: there is not enough rain to make any damage. It is composed by a 13 ticks sequence of 2/3 ticks followed by 2 ticks that can assume one of the value 0,1,2,3 and then a 0. Then, this event can be confused with a medium one (if not seen in its completeness).

As seen, these phenomena are not instantaneous, but they happen progressively in time, adding a given number of clouds on each tick until the phenomenon is completed.

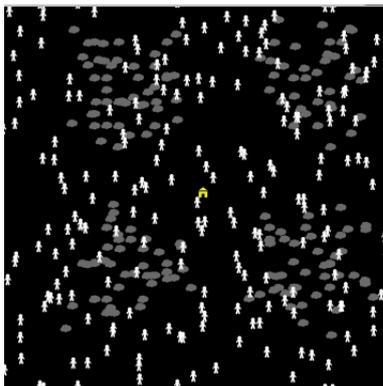


Figure 3: An example of a world after an event. Starting from the quadrants in the upper left and proceeding clockwise, we can see events 3, 1, 3 and 2.

The four quadrants are independent from each other but there can be an indirect influence as agents can have neighbors in other quadrants. In each quadrant, each event has a fixed probability to happen:

1. 10% for critical event;
2. 20% for medium event;
3. 70% for light event.

These events are also correlated to the alarms that the authority raises. In fact, as previously said, the authority is characterized by a standard deviation. We use it to produce the alarm generated by the authority and from it depends the correctness of the prediction. In particular, we considered four kinds of authorities: **reliable and strongly communicative**, **reliable and weakly communicative**, **non reliable and strongly communicative**, **non reliable and weakly communicative**.

### 3.4.1 Own Evaluation

How should agents evaluate the phenomena they see? We propose an empirical way to evaluate them, taking into account how phenomena are generated and how they can evolve.

Considering what we just said, agents can see a sequence of 3 clouds or a sequence of 2 and 3 clouds. The first one can lead to a critical or a medium event, the second one to a medium or light event.

Table1 provides a complete decryption of how agents evaluate what they see:

Programmatically, agents make a pattern matching of what they see, respecting the order of the table. As readers can notice, there are just a few cases in which agents are completely sure of what is going to happen.

Table 1: This table provides a comprehensive description of agents evaluation in each possible situation

| Event seen                                      | Probability of critical event | Probability of medium event    | Probability of light event |
|---|-------------------------------|--------------------------------|----------------------------|
| 15 ticks of 3                                   | 90%                           | 10%                            | 0%                         |
| 14 ticks of 3                                   | 80%                           | 20%                            | 0%                         |
| 13 ticks of 3                                   | 50%                           | 50%                            | 0%                         |
| 12 ticks of 3                                   | 10%                           | 80%                            | 10%                        |
| 11 ticks of 3                                   | 5%                            | 70%                            | 25%                        |
| 13 ticks of 3 and 1 tick of 2                   | 0%                            | 80%                            | 20%                        |
| 13 ticks of 3 followed by (2,2), (2,3) or (3,2) | 0%                            | 100% (full-blown) medium event | 0%                         |
| 15 ticks of 2 or 3                              | 0%                            | 90%                            | 10%                        |
| 14 ticks of 2 or 3                              | 0%                            | 80%                            | 20%                        |
| 13 ticks of 2 or 3                              | 0%                            | 50%                            | 50%                        |
| 12 ticks of 2 or 3                              | 0%                            | 20%                            | 80%                        |
| 11 ticks of 2 or 3                              | 0%                            | 10%                            | 90%                        |
| Any other case                                  | 0%                            | 0%                             | 100%                       |

### 3.5 Workflow

Each simulation is divided into two steps. The first one is called **"training phase"** and has the aim of letting agents make experience with their information sources, so that they can determine how much each source is reliable.

At the beginning of this phase, we generate a world containing an authority and a given number of agents, with different abilities in understanding weather phenomena.

At the time  $t_0$  the authority gives forecast for a future temporal window (composed by 16 ticks) including an alarm signal, reporting the level of criticality of the event that is going to happen in each quadrant (critic = 3, medium = 2, light =1). This information will reach each single agent with a probability given by the **authority communicativeness**.

Being just a forecast, it is not sure that it is really going to happen. It will have a probability linked to the precision of the authority (depending from standard deviation). However, as a forecast, it allows agents to

evaluate the situation in advance, before the possible event. Event in fact starts at  $t_1$  and, as previously said, lasts for 16 ticks.

During the decision making phase, agents check their own information sources, aggregating the single contributes according to the corresponding trust values. They estimate the possibility that each event happens and take the choice that minimizes the risk. Then, accordingly to their own decision-making deadlines, agents will choose how to behave.

While agents collect information they are considered as "thinking", meaning that they have not decided yet. When this phase reaches the deadline, agents have to make a decision, that cannot be changed anymore. This information is then available for the other agents (neighborhood), which can in turn exploit it for their decisions. At the end of the event, agents evaluate the performance of the source they used and adjust the corresponding trust values. If they haven't been reached by the authority, there will not be a feedback on trust but, as this source was not available when necessary, there will be a reduction of trust linked to the kind of event that happened: -0.15 for a critical event, -0.1 for a medium event, -0.05 for a light event.

This phase is repeated for 100 times (then there will be 100 events) so that agents can make enough experience to judge their sources.

After that, there is the textbf"testing phase". Here we want to understand how agents perform, once they know how much reliable their source are. In order to do that, we investigate how they perform in presence of a fixed map [3 1 3 2]. In this phase, we will compute the accuracy of their decision (1 if correct, 0 if wrong).

### 3.6 The Decision-making Phase

Once consulted all the three sources of information, agents subjectively estimate the probability that each single event happens:

1.  $P_{critical\_event}$  = probability that there is a critical event;
2.  $P_{medium\_event}$  = probability that there is a medium event;
3.  $P_{light\_event}$  = probability that there is a light event.

They will react according to the event that is considered more likely to happen. There are three possible choices:

1. Escape: agents abandon their homes.
2. Take measures: agents take some measure (quick repairs) to avoid possible damages due to weather event;
3. Ignore the problem: agents continue doing their activities, regardless of possible risks.

We assume that there is a time limit for taking a decision. This deadline is fixed to 15 ticks. Agents have to decide within this moment.

### 3.7 Platform's Input

The first thing that can be customized is the **number of agents** in the world. Then, one can set the value of the two parameters  $\alpha$  and  $\beta$ , used for the sources trust evaluation.

It is possible to change the **authority reliability**, modifying its standard deviation, and the **authority communicativeness**, that represents the probability that each single citizen will receive the message of the authority. Concerning the training phase, it is possible to change its **duration** and determine the **probability of the events** that are going to happen on each quadrant, while in the testing phase, that lasts just for 1 event, one can configure what we call the **event map**: it is the set of the four events relative to the four quadrants, starting from the one top left and proceeding clockwise.

## 4 Simulations

Once realized the platform, we decided to use it to investigate how different authority's behavior affects citizens choice and the trust they have in their information sources. We believe in fact that authority's choices affect not only citizens individually and directly, but also, by a social effect, citizens not directly reached by the authority. To do that, we investigated a series of scenarios, populated by equal populations, but in presence of different

authorities. Then we analyze how citizens respond to these changes, measuring their trust values and the choice they make in presence of possible risks.

Simulation results are mediated through 500 cases, in order to delete the variability that a single case can have.

#### 4.1 Simulation's results

Simulation settings:

1. **number of agents:** 200;
2.  **$\alpha$  and  $\beta$ :** respectively 0.9 and 0.1;
3. **authority reliability:** we used the value 0.3 to shape a very reliable authority (its forecast are correct about the 90% of time) and 0.9 to shape a non reliable authority (its forecast are correct about the 50% of time);
4. **authority communicativeness:** 100% for all the three events (strongly communicative), meaning that each agent will receive authority's forecast or 30% for all the three events (weakly communicative), meaning that each agent will receive authority's forecast only the 30% of the time;
5. **training phase duration:** 100 events;
6. **probability of the events:** 10% critical event, 20% medium event, 70% light event;
7. **event map:** [3 1 3 2].

To help understanding our results, we are going to show them together. Let us start from the trust analysis

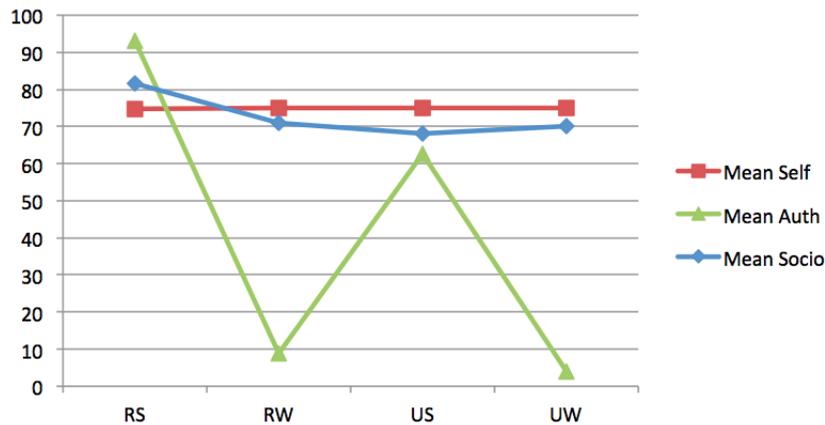


Figure 4: Here we represent the average trust value of all the agents on the three information sources in the four scenarios presented before.

Where RS = reliable strongly communicative, RW = reliable weakly communicative, US = unreliable strongly communicative, UW = unreliable weakly communicative.

Notice that, as implemented, 1/3 of citizens will quite always be able to quite completely understand the event, another 1/3 of them can understand quite well what is going to happen but it has less confidence on its evaluations, the remaining 1/3 of citizens is not a good evaluator. This fact stands for all the scenarios and that's why we have a standard value of average self trust.

Let's start analyzing results case by case. In this first **case RS** (reliable strongly communicative), the authority is reliable and all the agents have access to its information. This leads to a high level of authority trust, but also to a high level of social trust, as the information communicated at the social level is due to the one coming from the authority and the one directly seen by citizens.

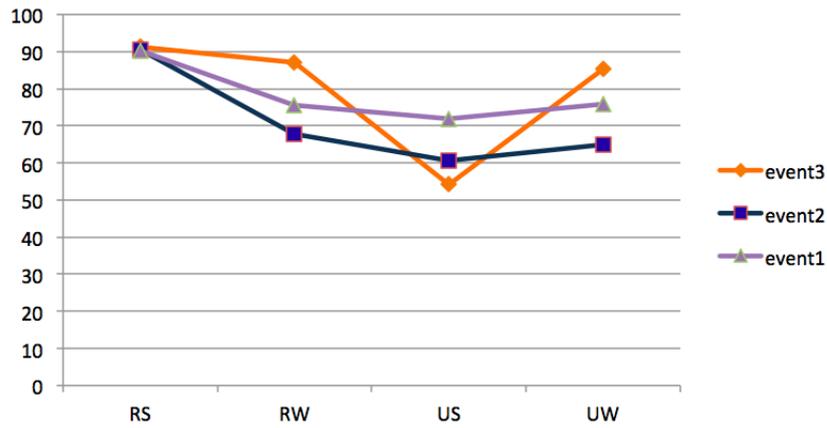


Figure 5: Here we represent the average performances of all the agents for the three events in presence of four different authorities.

In the **RW case** (reliable weakly communicative) we have a reliable authority, but it rarely communicates its information. This involves a lack of trust in that source as, even if it is reliable, it is quite always not available. Socially, this results in a lower level of social trust, as agents' decisions are just based on what they see (and just a part of them is able to directly understand the weather phenomenon). Even the agents' accuracy lowers down. Being the authority quite always unavailable, agents have to rely on their own abilities. In fact this effect is particularly strong for light events, as agents have less probability to read it correctly. Conversely, it is less strong on critical events, as is it easier to predict.

Let us see what happens when the authority is no more reliable, but it starts communicating wrong information. In the **US case** (unreliable strongly communicative), the average authority trust is 62.75%; this value comes from the fact that the authority reports correct information about the 50% of time. Remember that it is not true that wrong information is evaluated as 0. If it is not completely wrong (the authority predicted an event immediately near to the actual one) the evaluation will be 0.33. Most of the agents identify their own observation as the better information source (as the authority is not so much trustworthy). Then social decisions are mainly influenced by this component, but there is also a minimal influence of the authority performance. That is why we have a decrement on social trust.

The last case, the **UW** (unreliable weakly communicative), is supposed to be the worst one; the authority is not reliable and is weakly communicative. As we can see, the average authority trust is the lowest among the four cases, as even when available there is a good probability that the reported information will not be correct. We have a low value of average social trust, but it is higher than the third case. Again, because of an unavailable and inaccurate authority, agents will rely on their self. Let's then try to see the big picture, also comparing cases to each other.

In the first case (RS) we have the highest values of authority trust: it is a reliable available source, so that agents can rely on it. The authority trust has a good value also in the US case meaning that, in order to be trustworthy, it is important to be available to citizens, even if not always with correct information. Considering the RW and UW cases, they seem to be very similar. Here in fact regardless of authority's reliability, trust on the authority is very low. Even the average social trust seems to be the same in the RW and UW cases. It reaches its maximal point in the RS case, being the other two sources quite always right, and its minimal point in the US, when the authority reaches all the agents, but it spreads incorrect information.

Summarizing, the US case seems to be good from the authority's point of view, but it seems to have a negative social impact.

Taking into account performances, as expected the best case is the RS one; having just trustworthy sources, agents' performances are very high. Again the RW and UW cases, in which the authority is unavailable, are quite the same (actually the UW cases' values are a little bit lower) meaning that the if the authority is unavailable, it is no more important how much competent it is. The worst case is the US one. Here we have

that all the agents' performances decreases to their lowest value.

Notice that the event3 is the one who suffers more. To understand this phenomenon it necessary to take into account the table in section 2.5.1. Let us compare what happen in case of critical event and of light event. In case of critical event:

1. 1/3 of the population will estimate a 90% probability of critical event;
2. 1/3 of the population will estimate a 80% probability of critical event;
3. 1/3 of the population will estimate a 50% probability of critical event.

Conversely, in case of light event we have that:

1. 1/3 of the population will estimate a 100% probability of light event in the 75% of times, and a 10% probability of light event in the 25% of times;
2. 1/3 of the population will estimate a 100% probability of light event in the 50% of times, and a 20% probability of light event in the 50% of times;
3. 1/3 of the population will estimate a 50% probability of light event;

Then, even if it is more difficult for agents to detect a light event, when detected they will have a 100% certainty. On the contrary it is easier to them to identify a critical event, but not with high level of certainty. This means that, when there is the influence of another information sources that report wrong information, its influence on agents will be stronger in case of critical event rather than light event.

Finally, one could ask if it is better to have a reliable authority but not always available (RW) or an unreliable authority that has a strong presence (US). These results clearly state that the RW case is better, considering citizens' performance. This is due to the fact that, even if each individual citizen will receive right information from the authority about the 27% of the times in the RW case and about the 50% of the times in the US case, in the RW case the positive effect of the authority is widespread by the social effect. Then even if the authority is not does not reach everyone directly, it can count on the social effect to do it.

## 5 Conclusion

In this work we presented an articulated platform for social simulation, particularly suited for studying agents' choice in presence of critical weather phenomena. To this aim, we realized an ad-hoc Bayesian trust model, used by agents to evaluate and aggregate information coming from different information sources.

Using this framework, we were able to show some interesting results.

Through the training phase the agents learn to attribute to the different information sources the right values of trustworthiness and, as a consequence, they are able to perform quite effectively. In particular, two behaviors of the authorities are interesting: reliable and weakly communicative, not reliable and strongly communicative. They are a good simulation of the real cases in which the best prediction of a weather event is the more temporally close to the event itself (when becomes difficult to effectively spread the information: time for spreading is little). On the contrary, a prediction of a weather event can be effectively spread when there is big time for the spreading (far from the event), but this is in general a very inaccurate prediction.

Very interesting is the compensatory and integrative role of the social phenomenon (observation of the others' behavior) that guides the performances of the agents upwards when just one of the two other sources results as reliable.

## 6 Acknowledgements

This work is partially supported by the project CLARA-CLoud pLatform and smart underground imaging for natural Risk Assessment, funded by the Italian Ministry of Education, University and Research (MIUR-PON).

## References

- [Burnett et al, 2010] Burnett, C., Norman, T., and Sycara, K. 2010. Bootstrapping trust evaluations through stereotypes. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10). 241-248.
- [Burnett et al, 2013] Burnett, C., Norman, T., and Sycara, K. (2013) Stereotypical trust and bias in dynamic multiagent systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(2):26, 2013.
- [Castelfranchi and Falcone, 2010] Castelfranchi C., Falcone R., *Trust Theory: A Socio-Cognitive and Computational Model*, John Wiley and Sons, April 2010.
- [Conte and Paolucci, 2002] Conte R., and Paolucci M., 2002, *Reputation in artificial societies. Social beliefs for social order*. Boston: Kluwer Academic Publishers
- [Falcone and Castelfranchi, 2008] Falcone R, Castelfranchi C, (2008) Generalizing Trust: Inferencing Trustworthiness from Categories. In Proceedings, pp. 65 - 80. R. Falcone, S. K. Barber, J. Sabater-Mir, M. P. Singh (eds.). *Lecture Notes in Artificial Intelligence*, vol. 5396. Springer, 2008
- [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
- [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.
- [Liu et al, 2014] B. Liu, *Uncertainty theory 5th Edition*, Springer 2014.
- [Melaye and Demazeau, 2005] Melaye, D., & Demazeau, Y. (2005). Bayesian dynamic trust model. In *Multiagent systems and applications IV* (pp. 480-489). Springer Berlin Heidelberg.
- [Quercia et al, 2006] Quercia, D., Hailes, S., & Capra, L. (2006). B-trust: Bayesian trust framework for pervasive computing. In *Trust management* (pp. 298-312). Springer Berlin Heidelberg.
- [Sabater-Mir, 2003] Sabater-Mir, J. 2003. *Trust and reputation for agent societies*. Ph.D. thesis, Universitat Autònoma de Barcelona
- [Sabater-Mir and Sierra, 2001] Sabater-Mir J., Sierra C., (2001), Regret: a reputation model for gregarious societies. In *4th Workshop on Deception and Fraud in Agent Societies* (pp. 61-70). Montreal, Canada.
- [Wand and Vassileva, 2003] Wang, Y., & Vassileva, J. (2003, October). Bayesian network-based trust model. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on* (pp. 372-378). IEEE.
- [Wilensky, 1999] Wilensky, U. (1999). NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.
- [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).