When Agents Communicate Hypotheses in Critical Situations

Gauvain Bourgne[1], Nicolas Maudet[1], and Suzanne Pinson[1]

LAMSADE, Université Paris-Dauphine
Paris 75775 Cedex 16 (France)
Email: {bourgne,maudet,pinson}@lamsade.dauphine.fr

Abstract. This paper discusses the problem of *efficient propagation of uncertain information* in dynamic environments and critical situations. When a number of (distributed) agents have only partial access to information, the explanation(s) and conclusion(s) they can draw from their observations are inevitably uncertain. In this context, the efficient propagation of information is concerned with two interrelated aspects: spreading the information as quickly as possible, and refining the hypotheses at the same time. We describe a formal framework designed to investigate this class of problem, and we report on preliminary results and experiments using the described theory.

1 Introduction

Consider the following situation: witness of a threathening and unexpected event, say a fire in a building, Jeanne has to act promptly to both escape the danger and warn other people who might get caught in the same situation. However, there are no official signs or alarms indicating where the fire actually started: Given her partial knowledge of the situation, Jeanne may build some hypotheses explaining her observations (where the fire did start in the first place, maybe why), but the conclusions she may reach would remain *uncertain*. (That is, uncertainty here lies on the fact that she has incomplete knowledge of the world, rather than untrusted perceptions of this world). In addition, there is no way for Jeanne to trigger an alarm. In other words, Jeanne will try to both circulate the information in order to spread the information to colleagues, and refine the hypothesis at the same time. Typically, Jeanne faces two questions:

- What information should I transmit?
- To whom should I transmit this information?

Clearly, these two questions are interelated. Depending on the person Jeanne selected to communicate with, she may decide to transmit different messages: the objectives being to ensure that the transmitted information can be used efficiently in the next transmission, and so on. This defines, we believe, a problem of efficient propagation of uncertain information. The purpose of this paper is to put forward a formal framework expliciting both the reasoning and communicational aspects involved in these situations. We explore some preliminary properties of the proposed framework and interaction protocol, and illustrate our approach with a case study experimented using the described theory.

The remainder of this paper is as follows. Section 2 presents the formal reasoning machinery that we shall use in the framework: it heavily builds upon Poole's Theorist system [14]. Section 3 details the communication module, and explores specifically some properties of a protocol designed to exchange hypothesis. Section 4 describes our case study example, instantiating the proposed framework. The situation involves a number of agents trying to escape from a burning building. We give the detail of a simple example, showing how critical, in this crisis context, can be the decisions taken by agents as to whether/what communicate. Section 5 draws connections to related works, and Section 6 concludes.

2 Agents Reasoning

This section introduces the formal machinery involved in the agents reasoning process. The described situation suggests agents able to deal with partial perception of the world, to build hypotheses from observations they make, to draw conclusions from a set of explanations, and to communicate with each other in order to exchange pieces of information. Agents reasoning process builds on Poole's framework [14], which allows to elegantly combine both the explanation and the prediction processes, using a single axiomatization. By formulae we mean well-formed formulae in a standard first order language. Each agent is (slightly modified version) of an instance of a Theorist system [14]:

$$\langle \mathcal{F}, \mathcal{H}, C, O, E, \leq \rangle$$

where

- $\mathcal{F}$ a set of *facts*, closed formulae taken as being true in the domain
- $\mathcal{H}$ a set of formulae which act as *possible hypotheses*, common to all agents
- C a set of closed formulae taken as *constraints*, common to all agents
- O is a set of grounded formulae representing the *observations* made so far by the agent. Each agent believes every observation in this set to be true.
- E is the set of *preferred explanations*, it is the set of all justifiable explanations of the observation set O
- $\leq$ is the *preferrence relation*, a pre-order on the explanations common to all agents

We first recall a number of basic definitions.

Definition 1 (Scenario [14]). *A scenario of* $(\mathcal{F}, \mathcal{H})$ *is a set* $\theta \cup \mathcal{F}$ *where* θ *is a set of ground instances of elements of* $\mathcal{H}$ *such that* $\theta \cup \mathcal{F} \cup C$ *is consistent.*

In the following, we shall also refer to the conjunction h of the elements of θ as the *hypothesis* associated to this scenario.

Definition 2 (Explanation of a closed formulae [14]). *If g is a closed formula, then an explanation of g from $(\mathcal{F}, \mathcal{H})$ is a scenario of $(\mathcal{F}, \mathcal{H})$ that implies g.*

We now introduce a couple of further notions that proved to be appropriate in our context. Events occuring in the world and observed by the agents may or may not be explained, or contradicted, by the agent model.

Definition 3 (Positive observation). *A positive observation of $(\mathcal{F}, \mathcal{H})$ is an observation $o \in O$ such that there exists an explanation of o from $(\mathcal{F}, \mathcal{H})$*

Definition 4 (Negative observation). *A negative observation of $(\mathcal{F}, \mathcal{H})$ is an observation $o \in O$ such that there exists an explanation of $\neg o$ from$(\mathcal{F}, \mathcal{H})$*

In the following, we shall note $P(O)$ to refer to the set of all positive observations of $(\mathcal{F}, \mathcal{H})$, and $N(O)$ to refer to the set of all negative observations of $(\mathcal{F}, \mathcal{H})$. Note that this is not necessarily a partition: some observations may have no explanation, while some others may have both positive and negative explanations.

Definition 5 (Explanation of an observation set). *If O is a set of observations, an explanation of O from $(\mathcal{F}, \mathcal{H})$ is an explanation ξ of $P(O)$ such that $\xi \cup C \cup N(O)$ is consistent (which implies the consistency of $\xi \cup C \cup O$).*

Definition 6 (Justifiable explanation). *A justifiable explanation of O from $(\mathcal{F}, \mathcal{H})$ is an explanation such that if any element of its associated hypothesis set θ is removed from it, it is no longer an explanation of O.*

Based on this system, we also define, for each agent a_i:

1. H_i, the set of *preferred hypotheses* associated with E_i, the set of justifiable explanations. For a given set of observation O_i, $\mathcal{E}_{exp}$, the *explanation function* returns the set of all justifiable explanations of O_i from $(\mathcal{F}, \mathcal{H})$. $\mathcal{E}_{hyp}(O_i)$ gives the set of hypotheses associated with $\mathcal{E}_{exp}(O_i)$. We assume $\mathcal{E}_{exp}$ and $\mathcal{E}_{hyp}$ to be deterministic, and common to all agents.
2. h is the *favoured hypothesis* from E. The agent choses one favoured hypothesis among its own minimal hypothesis according to the preferrence relation.

In summary, for each agent we have:

- $E_i = \mathcal{E}_{exp}(O_i)$
- $H_i = \mathcal{E}_{hyp}(O_i)$
- $h_i \in min(H_i)$

This ensures that h_i is associated with a minimal justifiable explanation for O_i, that is :

- h_i is consistent with O_i, that is $\nexists o_i \in O_i$ s.t. $h_i \models \neg o_i$
- h_i explains all elements of $P(O_i)$

– h_i is justifiable from O_i, that is for each clause c_k of the conjunction h_i ($h_i = h'_i \wedge c_k$), there is an element o of $P(O_i)$ such that $h_i \models o$ but $h'_i \not\models o$.
– h_i is minimal according to the preorder $\leq$

Typically, as suggested by the aforementioned model, different explanations will exist for a given formula. What should be the preference relation between explanations? Clearly there can be many different ways to classify prefered explanations. In [14], different comparators are introduced. In our framework, we shall use variants of two of them:

1. *minimal explanation*— prefer the explanations that make the fewest (in terms of set inclusion) assumptions. In other words, no strict subset of a minimal explanation should also be an explanation.
2. *least presumptive explanation*— an explanation is less presumptive than another explanation if it makes fewer assumptions (in terms of what can be implied from this explanation together with the facts)

Now we need to see how these agents will evolve and interact in their environment. In our context, agents evolve in a dynamic environment, and we classicaly assume the following *system cycle*:

1. *Environment dynamics*: the environment evolves according to the defined rules of the system dynamics
2. *Perception step* : agents get perceptions from the environment. These perceptions are typically partial (*e.g.* the agent can only see a portion of the map), but we assume that they are *certain*, in the sense that the sensors are assumed perfect.
3. *Reasoning step*: agents compare perception with predictions, seek explanations for (potential) difference(s), refine their hypothesis, draw new conclusions. More precisely, during this step, if the agent perception prove its hypothesis false, the agent computes the possible explanations for these new perception, given its previous perception. It makes use of Theorist for this task. It must then select the action to be executed in the next phase.
4. *Action step*: agents modify the environment by executing the action selected by the previous deliberation steps.

What remains to be described, of course, is the interaction module and the way agents will exchange hypotheses and observations.

3 Agent Communication

In our system, observations are not only made directly by agents (by perceiving the environment): they can also result from communication between agents. The cycle is then augmented with an explicit *communication step*, which directly follows the *reasoning step*. During the *Communication step*, agents engage communication with other agents to warn of their observation and tune up their hypothesis. In a given round, a given agent can only communicate with *one* agent. If that agent is occupied talking to another agent, it must wait or choose a different agent to communicate with. We now describe the interaction protocol pictured in Fig. 1, together with agents' behaviour.

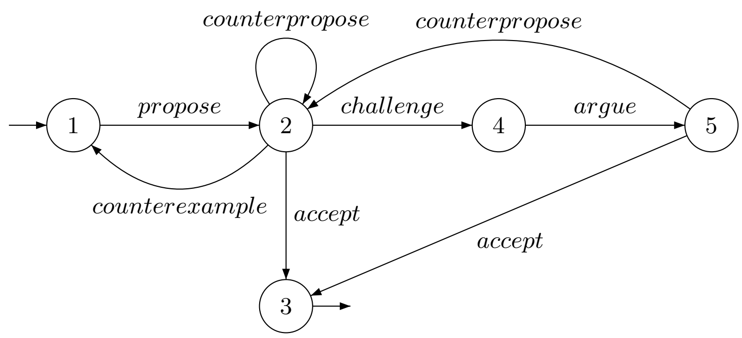


Fig. 1. Hypotheses Exchange Protocol.

3.1 Description of the Interaction Protocol and Strategies

Upon receiving a hypothesis h_1 (*propose*($h1$) or *counterpropose*($h1$)) from a_1, agent a_2 is in state 2 and has the following possible replies:

- if $\exists o_2 \in N(O2)$ s.t. $h_1 \models \neg o_2$, then the agent knows a counter-example that contradicts this hypothesis: he will communicate this counter-example and utter *counterexample*(o_2). We are back in state 1 of the protocol. Agent will then recompute his hypothesis with this new fact, and will propose h'_1.
- if $\exists o_2 \in P(O2)$ s.t. $h_1 \not\models o_2$, then the agent knows an example of positive observation that is *not* explained by this hypothesis: he will communicate this uncovered example and utter *counterexample*($o2$), as in the previous case.
- otherwise, no observation made by a_2 contradicts h_1 and h_1 implies $P(O_2)$, that h_1 is the hypothesis associated with an explanation of $O2$. We have then the following cases:
  - if the agent has no argument in favour of the hypothesis ($h1 \notin H_2$ where H_i is the set of the hypothesis associated to agent a_i's preferred explanations), he will *challenge* a_1 in order to obtain some arguments supporting this hypothesis. Agent a_1 is then bound to communicate an argument ($argue(arg)$)[1], leading to state 5. Upon receiving this argument, a_2 recomputes his hypothesis by using this argument. If h_1 is obtained, he will *accept*, leading to the final state 3. Otherwise, a different hypothesis h'_2 is obtained and proposed, leading back to state 2.

[1] Note that the agent keeps track of the communicated arguments, which allows him not to send twice the same argument to this agent during a communication step.

- otherwise $h_1 \in H_2$: h_1 is a hypothesis associated to a justifiable explanation of O. We have then two possibilities:
  - ∗ if h_1 is not preferred to h_2 in the sense of the defined preference relation, then agent a_2 would *counterpropose*(h_2), leading to state 2 with inverted roles.
  - ∗ otherwise, h_1 is necessarily prefered to h_2: a_2 will then respond *accept*, concluding the conversation (state 3).

3.2 Local Properties of the Interaction Protocol

We first investigate locally the properties of the proposed protocol, that is, the outcome of a single dialogue governed by the rules and decision process described in the previous subsection, and involving only two agents.

Lemma 1. *Let* $c = |O_1 \cup O_2| - |O_1 \cap O_2|$*. If* $c = 0$ *then* $O_1 = O_2$ *and* $H_1 = H_2$*.*

Proof. Clearly, $O_1 \cap O_2 \subseteq O_1 \cup O_2$. If $c = 0$, $|O_1 \cup O_2| = |O_1 \cap O_2|$, hence $O_1 \cup O_2 = O_1 \cap O_2$. Now because $O_1 \cap O_2 \subseteq O_1 \subseteq O_1 \cup O_2$, (and symetrically for O_2), we have $O_1 = O_2$. By virtue of the determinism of the explanation function, we conclude that $H1 = \mathcal{E}_{hyp}(O_1) = \mathcal{E}_{hyp}(O_2) = H_2$. □

The first property that needs to be verified is the *termination*. We show that this algorithm enjoys this property.

Property 1 (Termination). Termination is guaranteed, and the length of the interaction process (in terms of the number of exchanged messages) is bounded by $4 \times c + |O_1 \cap O_2|$.

Proof. Let $c = |O_1 \cup O_2| - |O_1 \cap O_2|$. By Lemma 1, we know that in case $c = 0$, it follows that $O_1 = O_2$ and $H_1 = H_2$ (in which case we note $O = O_1 = O_2$ and $H = H_1 = H_2$). Then observe that, $H = \mathcal{E}_{hyp}(O)$, together with the fact that $h_1, h_2 \in H$, guarantees that h_1 and h_2 are the favored hypotheses of the justifiable explanations of O. The following points then follow (i) $\nexists o \in O$ s.t. $h_1 \models \neg o$ or $h_2 \models \neg o$, (ii) $\nexists o \in P(O)$ s.t. $h_1 \not\models o$ or $h_2 \not\models o$, (iii) $h_1 \in H_2$ and $h_2 \in H_1$, and (iv) both $h_1, h_2 \in min(H)$, no hypothesis is then strictly prefered to the other one.

Given this, as soon as the system is in state 2, all termination conditions are met. But we also know that the message exchange between agents leads to state 2 every 3 messages at most. Termination is then guaranteed when $c = 0$.

We now need to prove that c will eventually reach the value 0. To do that, we will show that every 4 messages at most, it decreases of 1.

The first message leads to state 2. Without loss of generality, we assume that the last message is, say, from agent a_j to agent a_i (hypothesis h_j is then proposed to a_i). Following the agent's decision algorithm previously described, there are now four possibilities:

(i) $\exists o_i \in O_i$, s.t. $h_j \models \neg o_i$ or $\exists o_i \in P(O_i)$, s.t. $h_j \not\models o_i$, then a_i sends a counterexample o_i to a_j. In this case, $O'_j = O_j \cup \{o_i\}$ with $o_i \in O_i$ and $o_i \notin O_j$, which means that $|O'_j \cap O_i| = |O_j \cap O_i| + 1$, and $|O_j \cup O_i|$ remains unchanged. It follows that c is decreased by 1.

(ii) $\nexists o_i \in O_i$ s.t. $h_j \models \neg o_i$ and $\forall o_i \in P(O_i)$, $h_j \models o_i$ and $h_j \notin H_i$, then a_i requires an argument and a_j provides o_j. In this case, $O'_i = O_i \cup \{o_j\}$. If $o_j \in O_i$, then a_i repeats its challenge until he gets an observation o_j he didn't know before. Since a_j keeps track of its messages, at most $|O_1 \cap O_2|$ such messages can be exchanged. We eventually reach o'_j such that $O'_i = O_i \cup \{oj'\}$ where $o'_j \in O_j$ and $o'_j \notin O_i$.

(iii) $\nexists o_i \in O_i$ s.t. $o_i \models \neg h_j$ and $\forall o_i \in P(O_i)$, $h_j \models o_i$ and $h_j \in H_i$ but $h_j \notin min(H_i)$, then a_i respond with *counterpropose*(h_i). We are back in state 2, but now we are sure that $h_i \notin H_j$ (because $h_i \leq h_j$ and $h_j \in min(H_j)$, by definition), which means that we would be in case (i) or (ii).

(iv) $\nexists o_i \in O_i$ s.t. $o_i \models \neg h_j$, and $\forall o_i \in P(O_i)$, $h_j \models o_i$, and $h_j \in min(H_i)$, but then a_i accepts and the protocol terminates.

□

Corollary 1. *After termination, the following properties are guaranteed:*

- a_1 *and* a_2 *are consistent*
- a_1 *and* a_2 *have a hypothesis that explains both* $P(O_1)$ *and* $P(O_2)$
- a_1 *and* a_2 *have a hypothesis that is justifiable from* O_1 *and* O_2
- a_1 *and* a_2 *have a hypothesis that is minimal for* O_1 *and* O_2 *(that is* $h_1 \in min(\mathcal{E}_{hyp}(O_2))$ *and* $h_2 \in min(\mathcal{E}_{hyp}(O_1))$*)*

3.3 Global Properties of the Communication Protocol

The properties previously described hold locally, when only two agents interacts over one communication step. The next question is then to ask whether these properties can be guaranteed at a more global level. Clearly, many properties will not hold any longer when considered globally. One simple such property is the consistance, which cannot be transitive when only based on the bilateral hypothesis exchange protocol described. This can be observed by constructing an example where an agent a would first communicate a hypothesis to agent b, not revealing the full arguments supporting its position though. Now if b communicates in turn with a third agent, say c, it is clear that he may not be in a position to effectively defend this hypothesis, and may accepting c's hypothesis. a and c would then not be consistent. This is formally stated as follows.

Property 2. The consistance property guaranteed by the communication protocol is not transitive.

Proof. We construct the following counterexample : agent a_1 can communicate with agent a_2 and a_3, but agents a_2 and a_3 cannot communicate with each other. We assume that they share the following facts $\{p(X) \rightarrow r(X), q(X) \rightarrow$

$r(X), p(X) \rightarrow s(X)\}$, where $p(X)$ and $q(X)$ are hypothesis. We start with the following sets of observations $P_1 = \{r(a), \neg p(X)\}$, $P_2 = \{\ \}$, and $P_3 = \{s(a)\}$. Agent a_1 communicates $q(A)$, which is challenged by a_2. a_1 then provides an explanation $(r(a))$. Now a_2 communicates with a_3 and proposes $q(a)$, but a_3 has an additional observation, namely $s(a)$. Upon receiving this hypothesis, a_3 challenges a_2 and a_2 provides the only argument he has in possession: $r(a)$. But a_3 knows the further observation that $s(a)$ which makes the hypothesis $p(a)$ prefered. a_3 makes this counterproposal, a_2 challenges and a_3 gives his argument $(s(a))$. Now a_2 will accept. At this point of the interaction though, a_1 holds $q(a)$ as favoured hypothesis, while a_3 prefers $p(a)$, which is not consistent with $\neg p(X) \in P_1$. □

What this suggests is that we will need much more elaborated synchronization techniques to guarantee that these desirable properties still hold at the global level. However, in our context where time is a critical factor, and where communication can be highly restricted, it will be interesting to investigate in which situations simple protocols, like the one described here, can still give promising result and ensure an average good efficiency of the information propagation. As a first step towards this objective, we give in the next section an instance of the proposed framework and show a critical situation where communication and hypothesis exchange proves to be efficient.

4 A Case Study: Crisis Management

This section presents an instance of the general framework introduced earlier. We first describe the different parameters used to instantiate the framework. A complete example is then detailed.

4.1 Description of the situation

This experiment involves agents trying to escape from a burning building. The environment is described as a spatial grid with a set of walls and (thankfully) some exits. Time and space are considered discrete. Time is divided in rounds.

Agents are localised by their position on the spatial grid. These agents can move and communicate with other agents. In a round, an agent can move of one cell in any of the four cardinal directions, provided it is not blocked by a wall. In this application, agents communicate with any other agent (but, recall, a single one) given that this agent is in view, and that they have not yet exchanged their current favored hypothesis. Note that this spatial constraint on agents' communication could be relaxed in other contexts (which would require, in turn, to apply a more elaborated recipient choice algorithm).

At time t_0, a fire erupts in theses premises. From this moment, the fire propagates. Each round, for each cases where there is fire, the fire propagates in the four directions. However, the fire cannot propagate through a wall. If the fire propagates in a case where an agent is positionned, that agents burns and is considered dead. It can of course no longer move nor communicate. If an agent gets to an exit, it is considered saved, and can no longer be burned. It still can communicate, but need not move.

Agents know the environment and the rules governing the dynamics of this environment, that is, they know the map as well as the rules of fire propagation previously described. They also locally perceive this environment, but cannot see further than 3 cases away, in any direction. Walls also block the line of view, preventing agents from seeing behind them. Within their sight, they can see other agents and whether or not the cases they see are on fire. All these perceptions are memorised.

In order to deliberate, agents maintain a list of their possible explanations E (and a list of associated hypotheses H) explaining their observations about fire, and a prediction of fire propagation based on their favoured hypothesis h. The preference relation ($\leq$) is the following:

- the agent prefers the minimal explanation, taking into account only fire origins. In other words, an agent will prefer an explanation using an unique fire origin propagating over one using several sources.
- the agent prefer the least presumptive explanation, taking into account propagation and origins. In effect it means that the agent will favor an explanation considering the fire origin as closer to the observed manifestation.

Based on the reasoning described above, agents also maintain a list of possible escape route, sorted by simply favouring the shortest paths to exits.

4.2 Sample of Agents Theories

We now give a snapshot of the declarative representation of agents' knowledge, illustrating the different kind of rules involved in this example.

- Facts ($\mathcal{F}$) allow to represent the static elements of the environment, as well as the rules governing the dynamic of the environment. For instance, the following three rule state that there is indeed a vertical wall at location (0,1), that the fire can always be assumed to have started at the location it is observed, and eventually that the fire should propagate in four possible directions. This last one is an example of a rule justified in normal circumstances, but which may suffer exceptions: it is then represented as a default rule.

```
fact vwall(at(0,1)).
fact fire(T,at(X,Y)) <- origin(T,at(X,Y)).
default rule_propagates_L(T2,from(X2,Y)): fire(T,at(X,Y)) <-
        previous(X,X2), previous(T,T2), fire(T2,at(X2,Y)).
```

- The possible hypotheses set ($\mathcal{H}$), in this example application, is the set of all conjunctions of possible fire origin(s).
- Constraints (C) prevent default rules from applying. For example, the landscape includes walls and doors which prevent the fire from propagating.

```
constraint not rule_propagates_L(T,from(X,Y)) <- vwall(at(X,Y))
```

– Observations (O) can either be of the form `fire(T,at(X,Y))`, or of the form `nofire(T,at(X,Y))`

4.3 Example

We are now in a position to describe the steps of our illustrative example.

[Round t=0] A fire erupts at (6,6), but nobody can initially see it. It will propagate until t=3 before beeing seen.

[Round t=3]

Perception step. Agent a_1 sees fire at (3,6) (not expected), and agent a_3. Agent a_2 sees fire at (6,3) and (5,4) (not expected). Agent a_3 sees a_1.

Explanation step (a_1). Having computed an explanation for `fire(t=3, at (3,6))`, a_1 gets 12 possible explanations, each one exhibiting a single origin. One such explanation, as provided by the Theorist system, states that the fire may have started at location (4,5), before propagating to the north (i.e. from south) and to the west.

```
Answer is fire(t3, at(3, 6))
Theory is
[rule_propagates_R(t2, from(4, 6)),
rule_propagates_D(t1, from(4, 5)),
origin(t1, at(4, 5))]
```

To classify these hypothesis, he first selects the minimal hypothesis considering only the origin. In this case, all the hypothesis suppose only one origin for the observed fire. Among those, he then selects the less presumptive hypothesis. In this case, the selected hypothesis is:

```
[origin(t3, at(3, 6))]
```

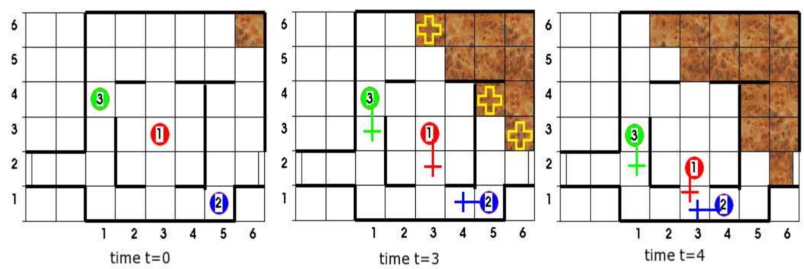
Explanation step (a_2)**.** Searching explanations for fire at (6,3) and (5,4), a_2 gets 6*6 possible explanations, such as :

```
Answer is fire(t3, at(6, 3)) and fire(t3, at(5, 4))
Theory is
[rule_propagates_R(t2, from(6, 4)),
origin(t2, at(6, 4)),
origin(t3, at(6, 3))]
```

Among those theories, only four of the explanations propose a common origin, and as such are minimal according to the origin criteria. Among those four, the less pre-emptive one is eventually:

```
[rule_propagates_R(t2, from(6, 4)),
rule_propagates_D(t2, from(6, 4)),
origin(t2, at(6, 4))].
```

Communication step. Agents a_1 and a_3 are the only agent seeing each other. Agent a_3 has no reason to initiate a communication, but a_1 has one: it has just changed its hypothesis and will try propagating and validating it. a_3 asks for arguments and a_1 sends it `fire(t=3,at(3,6))`. With this facts, Agent 3 recomputes its hypothesis and get the same favoured hypothesis. The hypothesis is confirmed and the communication stopped.



[Round t=4]

Action step. a_3 moves towards the west exit, which is the closest exit. a_1 moves towards the east exit, for the same reason. Although it is closer to the east exit, a_2 moves towards the west exit because it predicts that fire will arrive at the east exit before it can go out this way.

Perception step. Agent a_1 sees a_2 and conversely. All the fire seen by agents were predicted during this step.

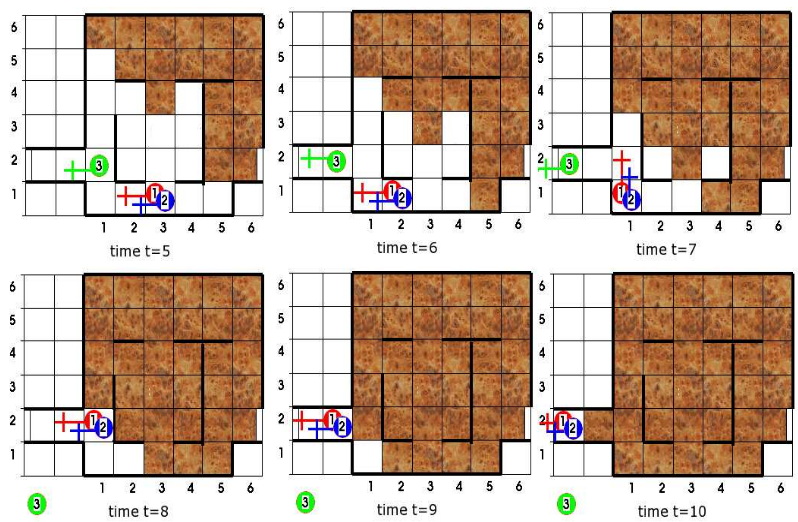
Explanation step. No agents has been confronted to unpredicted events. They have no need for explanation and just trim their hypothesis list.

Communication step. Agents a_1 and a_2 will communicate. Agent a_1 sends its hypothesis (`origin(t=3,at(3,6))`). As this hypothesis is not invalidated by its perception but does not belong to its hypothesis list, a_2 asks for arguments. Agent a_1 sends argument (`fire(t=3,at (3,6))`), and a_2 then computes possible explanations for this and its perception, and gets 6*6*12 possible explanations. Among those, only one contains a common origin for the three observed fires:

```
[rule_propagates_R(t2, from(6, 4)),
rule_propagates_D(t2, from(6, 4)),
rule_propagates_D(t1, from(6, 5)),
rule_propagates_D(t0, from(6, 6)),
rule_propagates_R(t3, from(4, 6)),
rule_propagates_R(t1, from(5, 6)),
rule_propagates_R(t0, from(6, 6)),
origin(t0, at(6, 6))].
```

Agent a_2 then proposes this hypothesis to a_1, which in turn ask for arguments. Finally both agreed upon this hypothesis.

Action step. Agent a_3 continues its escape towards the west exit. Agent a_2 confirms its chosen path with its new hypothesis, and keeps going towards the west door. Agent a_1, however, using its new hypothesis, discover that its escape route is bad. It changes its course to go towards the west exit.



[Round t=5 to 10] From time t=5 to time t=10, agents a_1, a_2 and a_3 exit the building. Agents a_1 and a_2 are closely followed by the fire: one false move would have been fatal! If a_1 did not communicate with a_2 or $_3$ it would not have been able to determine whether the fire was coming from left or right, and would have chosen the east exit and been trapped by the fire.

5 Related Work

Our approach has several facets that can be related to a number of related works. We now introduce some of these related works, starting with the studies of the notion of rumours in social science, that proved to be very inspiring for us.

Rumour in Social Sciences. Rumour is a complex phenomenon that has been the object of numerous studies in social science but is often seen as something that can only bring lies or diffamation. Studies of rumour in social science show, however, that there is more to rumour than just a routing or perception sharing system. Whereas the first studies, done during and after World War II, seem to consider rumour as something dangerous which should be avoided (rumours could lead to moral loss or information leak), more recent stances are somewhat more neutral or positive about it. J.N Kapferer [10] defines rumour as *"the emergence and circulation in the social body of information that either are not yet publicly confirmed by official sources or are denied by them"*. As an unofficial information, it must use alternative ways to be distributed, such as individual communication (gossip, word-of-mouth). He precises that a rumour spreads very quickly because it has value, and because this value decreases over time. Moreover the rightness of the content has no importance. A true rumour spread exactly like a false rumour. The exactitude of the content is not a criteria to define rumour. However, one can choose to take a slightly different perspective on the rumouring process. Shibutani [19] defines rumour as improvised news resulting from a collective discussion process, usually originating from an important and ambiguous event. In his own words, rumour is *"common use of the group individual ressources to get a satisfying intepretation of an event"*. In this case, the rumour is seen as being both an (i) information routing process and (ii) an interpretation and comment adding process. Crucially, the distorsion of information that is often seen as characteristic of rumour is seen as an evolution of the content due to continual interpretation by the group. A crucial aspect of rumour, of course, is that it is a decentralized process. The information propagates without any official control. It is deeply linked with spatial or communication constraint, and can be an efficient way to convey information in spite of these. It is also expected that this process is quite robust to agent error or disparition.

Distributed Diagnosis. The problem of multiagent diagnosis has been studied by Roos and colleagues [15, 16], where a number of distributed entities try to come up with a satisfying global diagnosis of the whole system. They show in particular that the number of messages required to establish this global diagnosis is bound to be prohibitive, unless the communication is enhanced with some suitable protocol. The main difference with our approach lies in the dynamic nature of our context, as well as in the constraints governing agents' interactions that we assume.

Argument-based Interaction. The idea of enhancing communication between agents by adding extra-information that may have the form of arguments has been influential over the last past years in the multiagent community [13]. However, although this approach has several clear advantages (*e.g.* improving expressivity, or facilitating conformance checking), its effectiveness regarding the speed and likelihood of fullfillment of the goal of the interaction has seldom been tested (exceptions are the work of [9], or [11], for instance).

Gossip Problem. Rumours and gossip first appeared within the distributed system community with the *gossip problem*: each agent has a distinct piece of information (called a rumour) to start with. The goal is to make every agent know all the rumours [18]. Some variation of it are the *rumour-spreading problem*, where the agent to communicate to is selected each round by an adversary [1], and the *collect problem.* In the last one, each of n processes in a shared memory system have several pieces of information, and all these processes must learn all the values of all others while making as few as possible primitive read or write operations [17]. It has also been used for reaching consensus [6]. This differs from our approach, mainly because we do not seek to necessarily converge towards a common knowledge of (initially distributed) informations. Also, agents do not modify informations they propagate.

Gossip-based protocols. Each agent has a determined number of neighbours it can communicate with. Each time an agent receives a rumour, it transmits it to to a number of agents chosen at random among its neighbours. Then in turn, each of these agents would do the same. This rumour spreading is analogous to the spreading of an epidemic, which have been the object of mathematical studies [2] and can spread exponentionally fast. Such an information propagation system has first been used for replicated database consistency management [8]. It has been applied to unstructured peer-to-peer communities. Every time an agent detects a change in the system (that would be the rumour), it sends it to a random neighbour, and repeats this operation until it has contacted enough neighbour(s). Some anti-entropy mechanisms are sometimes used to ensure that every agent can get to know each change, even if the rumour has already died out [7]. Another application of these protocols is reliable multicast [3]. It aims at propagating an information from an agent to another agent without a centralised source or knowledge of the system topology, and with a lower cost than with a simple flooding. It is robust to agent deficiency, and very scalable. A variation of it uses weight to enhance the reliability in specific topology [12]. This approach is related to the "recipient selection" aspect of our problem. However, the transmitted information is, again, assumed to be unaffected by agents' reasoning.

Rumour routing. Another approach of rumour as an alternative to flooding is *rumour routing* [4]. In the context of sensor networks, there is a need to transmit queries to agents having observed an event. A fast route between an agent making a request and the agents observing the events might be needed. It can be found by flooding event notifications or queries, and creating a network-wide gradient field [20], but it is a costly approach. Braginsky and Estrin instead propose to use a kind of traceable rumour. Each time an agent observes a new event, it sends an event notification rumour to a random neighbour. This neighbour transmits it in turn to another neighbour, keeping trace of whom it received it from, and how many agent(s) have acted as relay(s), creating rumour paths. When an agent needs to make a query, it sends it to one of its neighbours. If it has heard of the event concerned before, it transmits the query to the agent who told it the rumour, else it transmits to a random neighbour. Eventually, the query will cross the rumour path and be led to the right source. As in the preceding cases, rumour routing propagates pure information, therefore the main studied aspects are the velocity and robustness of these processes.

Reputation Systems. Buchegger and Le Boudec, for instance, use the term of rumour in a reputation system [5]. Their agents can make decisions about the reliability of others agents according to their previously observed behaviour, but also according to what others agents tell about it. In this case, rumour is primarily intended to mean "second-hand information". In this case, agents can keep track of previous partners' behaviours, and also report their observations to other agents. However, these agents are not able to explicitly reason over the justifications governing their decisions.

6 Conclusion

This paper discusses the problem of *efficient propagation of uncertain information* in dynamic environments and critical situations. When a number of (distributed) agents have only partial access to information, the explanation(s) and conclusion(s) they can draw from their observations are inevitably uncertain. In this context, the efficient propagation of information is concerned with two interrelated aspects: spreading the information as quickly as possible, and refining the hypothesis at the same time. We describe a formal framework designed to investigate this class of problem, and propose a simple protocol allowing hypothesis exchange. We also prove some preliminary properties of the protocol and report on an experiment conducted using the described theory.

An obvious advantage of this process (that we observed on the described example) is that agents do not wait to collect all data before providing and propagating hypotheses. In our example this allows agents to escape a building before being caught by the fire. When exactly temporary hypotheses are good enough to be acted upon is to be determined, but this process definitely enable quicker reaction to events than a static centralized data analysis.

The problem is that, of course, it can give incomplete or wrong hypothesis, as the very preliminary analysis of the global properties of the framework suggests. More elaborated communication techniques may then be investigated, allowing agents to backtrack and further refine their hypotheses. In critical situations however, it is unlikely that agents will dispose of sufficient resources to fully synchronize their hypotheses and observations. In consequence, we believe the situations as the one described in our case study to be well suited to such an approach. Further studies are required, however, to determine when exactly this kind of communication would be beneficial, but we expect quickly evolving systems to provide interesting applications. Whereas this paper has mainly focused on agents' reasoning and content selection, we plan to investigate in future research the related problem of recipient selection. Finally, it would also be interesting to consider more complex cases, for instance where agents may have unreliable perceptions of the world, or where malicious propagators of information could adopt an uncooperative behaviour.

Acknowledgments. We would like to thank the anonymous reviewers whose detailed comments helped to greatly improve the paper.

References

1. J. Aspnes and W. Hurwood. Spreading rumors rapidly despite an adversary. In *Proc. 15th ACM Symposium on Principles of Distributed Computing*, pages 143–151, 1996.
2. N. Bailey. *The Mathematical Theory of Infectious Diseases.* Charles Griffin and Company, London, 1975.
3. K. Birman, M. Hayden, O. Ozkasap, Z. Xiao, M. Budiu, and Y. Minsky. Bimodal multicast. *ACM Transactions on Computer Systems*, 17(2), 1999.
4. D. Braginsky and D. Estrin. Rumor routing algorithm for sensor networks. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, 2002.
5. S. Buchegger and J. Le Boudec. The effect of rumor spreading in reputation systems for mobile ad-hoc networks. In *Proc. of Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, 2003.
6. B. Chlebus and D. Kowalski. Gossiping to reach consensus. In *Proc., 14th ACM Symp. on Parallel Algorithms and Architectures*, 2002.
7. F. M. Cuenca-Acuna, C. Peery, R. P. Martin, and T. D. Nguyen. PlanetP: Using Gossiping to Build Content Addressable Peer-to-Peer Information Sharing Communities. In *Twelfth IEEE International Symposium on High Performance Distributed Computing (HPDC-12)*, pages 236–246. IEEE Press, June 2003.
8. A. Demers et al. Epidemic algorithms for replicated database maintenance. In *Proceedings of 6th ACM Symposium on Principles of Distributed Computing*, pages 1–12. Vancouver, British Columbia, Canada, 1987.
9. H. Jung and M. Tambe. Argumentation as distributed constraint satisfaction: Applications and results. In *Proceedings of AGENTS01)*, 2001.
10. J.-N. Kapferer. *Rumeurs, le plus vieux média du monde.* Points Actuel, 1990.
11. N. C. Karunatillake and N. R. Jennings. Is it worth arguing? In *Proceedings of First International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2004)*, pages 62–67, 2004.
12. M.-J. Lin and K. Marzullo. Directional gossip: Gossip in a wide area network. In *European Dependable Computing Conference*, pages 364–379, 1999.
13. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.
14. D. Poole. Explanation and prediction: An architecture for default and abductive reasoning. *Computational Intelligence*, 5(2):97–110, 1989.
15. N. Roos, A. ten Tije, and C. Witteveen. A protocol for multi-agent diagnosis with spatially distributed knowledge. In *Proceedings of AAMAS03*, pages 655–661, 2003.
16. N. Roos, A. ten Tije, and C. Witteveen. Reaching diagnostic agreement in multi-agent diagnosis. In *Proceedings of AAMAS04*, pages 1254–1255, 2004.
17. M. Saks, N. Shavit, and H.Woll. Optimal time randomized consensus - making resilient algorithms fast in practice. In *Proceedings of the 2nd ACM-SIAM Symposium on Discrete Algorithms*, pages 351–362, 1991.
18. S.Even and B. Monien. On the number of rounds needed to disseminate information. In *Proc. of the First Annual ACM Symposium on Parallel Algorithms and Architectures*, 1989.
19. T. Shibutani. *Improvised News : A Sociological Study of Rumor.* Indianapolis and New york, 1966.
20. F. Ye, G. Zhong, S. Lu, and L. Zhang. Gradient broadcast: A robust data delivery protocol for large scale sensor networks. *ACM Wireless Networks*, 2005.