# Distributed Belief Revision versus Distributed Truth Maintenance

Aldo Franco Dragoni, Paolo Giorgini and Paolo Puliti

Computer Sciences Institute
University of Ancona
Ancona, Italy, 60131

**Abstract**

*This paper outlines a distinction between Distributed Truth Maintenance and Distributed Belief Revision. The latter has a more complex conceptualization than the former and it needs the evaluation of special features as the relationship between the Informant's reliability and the Information's credibility. We point out some criteria to judge the qualities of a Distributed Belief Revision strategy from a global perspective. However, the performances of such a strategy can be estimated only on a simulation basis. The general framework for Assumption-Based Distributed Belief Revision that we present has been tested on our specific "CLUEDO" multi-agent simulation testbed. The model has been continuously improved during many cicles <modification of the model, simulation, evaluation of the performances> till the results presented in this paper.*

## 1 Introduction

### 1.1 Belief Revision versus Truth Maintenance

It is generally recognized that the abilities to detect contradictions, identify their culprits and readjust the knowledge base to remove them are important features to embed in an intelligent information system. This is what probably should be called "Consistency Maintenance" or "Truth Maintenance" (TM) [19] but has been often referred to as "Belief Revision" (BR) [10]. Maintaining consistency implies throwing part of the knowledge base and, obviously, we need good strategies to cut away the tumour limiting at the least (possibly at zero) the damage. Whether we move in an abstract and descriptive "default theory" scenario [17] or in a concrete and procedural Assumption Based Truth Maintenance framework (ATMS) [5], we have to reduce the contradictory knowledge space, to an "extension" in the former case or to a "context" in the latter one. Now, given an inconsistent knowledge base, there are many different extensions/contexts (too much in the worst case [12]) as possible solutions. In our view, BR can be defined as TM plus the selection of one (or more) of these solutions as the preferred one(s). This decision problem has been scarcely addressed ([10] p.31) mainly because it has been regarded as a domain dependent one. However, this problem is similar to that of selecting the best abductive explanation [2] or the most plausible diagnosis in model-based theories [7], and some general domain independent criteria to deal with these problems have already been presented [11] [6].

### 1.2 BR in a multi-agent world

In a forthcoming network of interacting Knowledge-Based Systems some nodes could join the network with low degrees of competence or non-cooperative intentions (may be destructive ones). In these cases we need good strategies to protect the overall Global Information Agency from the introduction of information pollution, making it able, as far as possible, to detect its unreliable member agents.

In such a realistic multi-agent scenario it becomes necessary to enlarge the classical idea of BR. In fact, to detect contradictions and identify their sources it is sufficient to maintain information about *what* has been told; but to properly "solve" a contradiction it is necessary to keep information about *who* said it or, in general, about where did that knowledge come from![1] We can take as certain the fact that an agent gave some information, but we can take the given information only as a revisable *assumption*. The BR system cannot leave the sources of the information out of consideration because of their relevance in giving the additional notion of "strength of belief" [9]. The *reliability* of the source affects the *credibility* of the information and vice-versa. It is necessary to develop systems that deal with couples

[1] it could also came from certain or hypothetical local information

<information, informant>.

### 1.3 Distributed BR versus Distributed TM

In a multi-agent environment, agents exchange knowledge and then make inferences based both on locally found and exchanged knowledge. The global BR task becomes problematic since agents must compute their beliefs locally, based also on beliefs communicated and justified externally. Recently, researchers from the Distributed Artificial Intelligence community [1] conceived distributed versions of the tms'paradigm [13] and of the atms'one [15], but still remain some terminological or conceptual confusions between Distributed tm and Distributed BR [14]. Distributed tm studies how local algorithms to maintain consistency will affect the consistency of the global distributed knowledge base, that is how to achieve global consistency from a local perspective. Distributed BR should study how these local TM algorithms plus some local criteria to choose the preferred set of beliefs and some policies of communication could affect each agent's opinions regard what is most credible and who is most reliable. With Distributed BR special questions arise that regard the global emergent epistemic behaviour; for instance:

a) does the proposed local BR strategies assure the various agents to converge gradually toward the same knowledge space?

b) if this global beliefs convergence is assured, is it stable and how much time will be necessary to achieve it?

c) is it possible for the various agents to detect those among themselves which are particularly unreliable?

d) if that is possible to what extent? i.e., what happens if most of the agents are largely unreliable?

e) is the overall global agency reliable? i.e., does it converge toward the more credible knowledge or not?, etc..

In this paper we propose an atms-based single-agent BR architecture, discussing the criteria to select beliefs from a single agent perspective. Then we introduce the metaphor of an investigation agency presenting CLUEDO, a multia-gent simulation testbed inspired by a common society game with which we are studying the properties of the proposed BR strategies from a global point of view.

## 2 A model for BR in a multi-agent environment

Here we present a BR model for a *single* agent exchanging knowledge with other companions. This model brings together assumption-based reasoning and special techniques to deal with uncertainty that distinguish between the assumption's credibility and the informant's reliability. The ATMS-based algorithm guarantees the "stability,"[1] "well-foundedness,"[2] and "logical consistency"[3] of the agent's knowledge base. Uncertainty management techniques will determine the more plausible set of beliefs for the agent to reason with. The model's basic architecture is sketched in Fig 1.

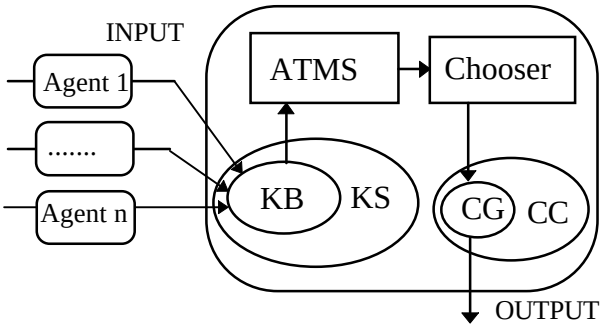


Fig 1. The basic architecture for Belief Revision

This ATMS doesn't distinguish between data and justifications; both appear in nodes as sentences of a decidable first order language (the latter as Horn clauses). There are two kinds of nodes: those introduced as *assumptions* and those deductively *derived* by a theorem prover as logical consequences of (the sentences in) other nodes. We call *Knowledge Base* (KB) the set of all the assumptions' nodes currently introduced, and we call *Knowledge Space* (KS) KB itself augmented with all the nodes currently derived from those in KB. No sentence will ever be removed from KS. Each node's *Origin Set* (OS) records the assumption nodes upon which it really ultimately depends (as derived by the theorem prover). The OS of an assumption node contains only the identifier of the node itself. A same sentence can appear in multiple nodes with different OSs.[4] The ATMS detects and stores in tables the *nogoods*; which are the minimally inconsistent subsets of KB. A subset of KB is inconsistent iff it is a superset of at least a nogood. A *good* is a subset of KB such that:

1. it is not inconsistent (it is not a superset of a nogood),
2. if augmented with whatever else assumption in KB it becomes inconsistent.

If, because of complexity reasons, the ATMS is not assured to find all the nogoods, then we substitute the notions of "weak-consistency" and "strong-

---

1 each element that has a valid justification is believed, while each element that lacks a valid justification is disbelieved.

2 there are no mutually dependent elements.

3 as far as is currently known to be inconsistent.

4 in particular, the same sentence can appear in an assumption node and in derived nodes at the same time.

inconsistency" to those of "consistency" and "inconsistency" [8]: a set of sentences is strongly-inconsistent if it has been proved to be inconsistent, and it is weakly-consistent if it hasn't been showed to be inconsistent but the search was not complete.

Given a KB, there is a bijective mapping between sets of nogoods and sets of goods. Each good has a corresponding context, which is the subset of KS made of all the nodes whose OS is subset of the good. The same nodes can belong to different contexts. The ATMS's ability to manage multiple contexts is very appealing from the BR point of view because it makes possible to compare the credibility of different maximally consistent sets of beliefs as a whole rather than to compare the credibility of different single beliefs.

A main feature of this BR model is the introduction of criteria to select the best context to reason with among the many possible outcomes of the ATMS. It is not the case to select which belief has to be thrown away to remove the contradiction, but, quite more generally, to choose which is the new preferred *good* among them in KB; this is the task of the Chooser. In ATMS-based reasoning the emphasis is placed on assumptions, so we do not care about the credibilities of derived sentences.[1] The BR criteria select the more plausible contexts by comparing the credibilities of their goods. We call *current good* (CG) the particular good chosen as the preferred one and *current context* (CC) its corresponding context. In a multi-agent environment, to choose CG it will be necessary to develop systems that deal with couples <information, informant>, evaluating in the same time the source's reliability and the information's credibility. We need adequate functions to model these relationships. To cope with the complexity of the matter we introduce three dynamically related parameters:

1.$r_s$, *reliability* of the source *s* estimated by the receiving agent,

2.$c_\alpha$, *credibility* of the assumption $\alpha$ estimated by the receiving agent,

3.$c_{\alpha,s}$, *source credibility* (*s-credibility*) of the assumption $\alpha$ estimated by the source *s*.

These parameters range from -1 to +1. A source with $r_s$=-1 is absolutely mendacious, a source with $r_s$=0 is unreliable and a source with $r_s$=1 is absolutely reliable. An assumption $\alpha$ with $c_\alpha$=-1 is absolutely incredible, an assumption with $c_\alpha$=0 is uncertain and an assumption with $c_\alpha$=1 is certain. A node has the following structure:

<*Identifier, Sentence,* OS, *Source, Credibility*>

The sources' current reliabilities and the assumptions's s-credibilities are collected in two global tables. The same sentence can appear in multiple nodes with different credibilities.[2] The credibility of a good is defined to be the average of the credibilities of the assumptions in it. The preferred context CC is chosen as the one associated to the most credible good. The intended meaning for CC is to be the maximal and globally most believable piece of knowledge currently available to the reasoning agent. A sentence is believed if and only if it appears in CC with a positive credibility. There are no relationships between the credibility of an assumption and that of its negated. However, although a contradictory set of assumptions is not incredible (its credibility is generally different from -1) the ATMS removes it. Generally, given an assumption $\alpha$, $c_{\alpha,s}$ and $c_\alpha$ are different because each receiving agent judges $c_\alpha$ from its point of view, that is from, at least, the following items.

a) Currently locally estimated reliability $r_s$ of the source. When an agent receives an information with a s-credibility $c_{\alpha,s}$ from a source whose reliability is $r_s$, it estimates the information's credibility by adding two terms:

$$c_\alpha = r_s \cdot c_{\alpha,s} + |r - r_s| \cdot g \cdot c_{\alpha,s}$$

where r is the agent's own *auto-reliability*. In the first term, whether the information is given as credible or not ($c_{\alpha,s}$ positive or negative), its credibility will be as more uncertain as more unreliable is considered the source. If a source is considered mendacious then the credibility and the s-credibility have different signs. If a source is considered reliable then the credibility and the s-credibility have the same sign. The following table summarizes the qualitative behaviour of this first term.

| $r_s$ | $c_{\alpha,s}$ | $c_\alpha$ |
|---|---|---|
| + | ++ | + |
| 0 | ++ | 0 |
| - | ++ | - |
| +/0/- | 0 | 0 |
| + | -- | - |
| 0 | -- | 0 |
| - | -- | + |

The second term introduces the parameter *g* that is the percentage of information received from the source s that already belongs to the receiving agent's CG with the same credibility sign. With this term we increase the credibility of a by a quantity proportional to *g* and to the distance between the auto-reliability *r* and the source's reliability $r_s$.

b) Local consistency with all the other assumptions in

[1] However, the credibility of a derived sentence should be that of the least credible assumption in its OS.

[2] In particular, the same source can provide the same Information at different times with different credibilities.

its KB. The discovery of a nogood affects the internal credibilities of its assumptions. Consider the nogood $\{\alpha, \neg\alpha\}$. If $c_\alpha$ and $c_{\neg\alpha}$ have different signs then their absolute values should increase; if they have the same sign and comparable absolute values, then their absolute values will decrease; if they have the same sign and very different absolute values then their minor credibility could change its sign; in the following table $c_\alpha$ and $c'_\alpha$ are the credibilities of $\alpha$ before and after the discovery of the nogood.

| $c_\alpha$ | $c_{\neg\alpha}$ | $c'_\alpha$ | $c'_{\neg\alpha}$ |
|---|---|---|---|
| -- | ++ | --- | +++ |
| ++ | -- | +++ | --- |
| ++ | ++ | + | + |
| -- | -- | - | - |
| ++ | + | + | - |
| -- | - | - | + |

The following function fits this qualitative behavior:

$$c'_a = c_a - \rho \frac{c_{\neg a}}{|c_a| + |c_{\neg a}|}$$

in which $c'_\alpha$ is bounded into the range $-1 \le c'_\alpha \le 1$. The parameter $\rho$ $(0 < \rho \le 1)$ take care of information cardinality:

$$\rho = \frac{N_c}{N_c + N_{nc}}$$

$N_C$: the number of assumptions that are into the biggest CC to which $\alpha$ belongs.
$N_{NC}$: the number of assumptions that are into the biggest nogood to which $\alpha$ belongs.

Generally, a nogood involves more than two assumptions and the same assumption can be involved in more than one nogood. In this case the new credibility of an assumption depends on the credibilities of all the other assumptions in the nogood for all the nogoods. The following function generalizes the preceding one:

$$c'_\alpha = c_\alpha - \rho \cdot \sum_{ng \in NG} \frac{C_{ng}}{|c_\alpha| + |C_{ng}|}$$

where:

$$C_{ng} = \frac{\sum_{(k \in ng) \wedge (k \neq a)} c_k}{|ng| - 1}$$

where *NG* is the set of nogoods to which $\alpha$ belongs and $|ng|$ is the cardinality of *ng* . This function treats all the nogoods as they were detected simultaneously.

These changes in the internal credibilities of the assumptions will affect their respective current source's reliability. This new reliability will be used to calculate the credibilities of the next information coming from that agent. The idea is simply that an Information source's reliability should decrease with the distance between the Information's credibility and s-credibility. This is an acceptable correspondence:

$$r_s = 1 - |c_\alpha - c_{\alpha,s}|$$

Given the set *R* of all the assumptions come from that source, the actual current source's reliability is the average of all the reliabilities for each assumption.

$$r_s = \frac{\sum_{\alpha \in R} 1 - |c_\alpha - c_{\alpha,s}|}{|R|}$$

Each agent calculates its own auto-reliability

$$r' = \frac{r + \frac{\sum_{a \in R} 1 - |c_\alpha - c_{\alpha,s}|}{|R|}}{2}$$

basing on the value r took before. In this way we keep the agent more conservative toward its own auto-reliability, i.e. it will be more confident in its own capabilities.

## 3 Distributed BR versus Distributed TM

When agents exchange beliefs and then make inferences based on them, then the concept of global TM becomes especially problematic. In [13], the authors' goal is that all the agents are both individually and mutually consistent with any other agent with whom they exchanged knowledge. We think that local consistency can be considered as a prerequisite, but any degree of global consistency should be considered only as a finishing post eventually to reach through adequate local BR strategies. Obviously, there is the risk that although the agents may come to an agreement, it may be the wrong consensus, since agents can update their beliefs with faulty information and then pass it on and contaminate other agent's beliefs.[1] It is unreasonable and intuitively wrong to impose the mutual consistency between communicating agents, it is better to let them stand by their beliefs based on their own view of the evidence. This permits the realistic possibility that nobody has uncompromised evidence or information. This is what [15] calls "Liberal Belief Revision Policy." In real world agents are not necessarily benevolent nor competent, so they can lye, deliberately or not. As far as the forthcoming Information Systems networks will spread over the world we'll need local BR policies which

[1] Davis and Smith's model of result sharing problem solving activity [18] runs this risk; they presume mutual credibility and trust among the agents.

resist the defilement of the information. We agree with [15] that there is no satisfactory answer to the question "How do we determine which agent is *right*?", but we can try to answer the questions "How do we determine which agent is *more reliable*?" and "How do we determine which information is *more credible*?"

Along with adequate local BR strategies, to reduce the risks of information pollution and/or monopoly we'll need also good local communication policies. Particularly, in our experimental sessions we make the following settlements:

1. sincere agents transmit not all their KS but only their preferred context CC,
2. agents *do not communicate the sources of the assumptions, but they present themselves as completely responsible for the knowledge they are passing on*; receiving agents consider the sending ones as the sources of all the assumptions they are receiving from them.[1]
3. agents *do not exchange opinions regarding the reliabilities of their companions*.

This general mechanism for decision is a mostly subconscious task to the agent, but nothing prevents us to consider agents aware of it so that they can conceive strategies to influence the others by means of communication. All agents basically use the same representation language (syntax and semantic) for beliefs. Communicating agents know what they are talking about, and they have a common "understanding" of the propositions they exchange."

### 3.1 A Simulation Testbed

Adopting the BR model presented in this paper as a local mechanism to manage and solve contradictions we hope to achieve:

1) the convergence of the agents' opinions regarding credibility and reliability;
2) this convergency's stability;
3) this convergency's correctness;
4) this convergency's robustness (up to what unreliability degree and how many unreliable agents we can extend the results?).

We remark that we hope to achieve these results without:
- the agents' comunicating information concerning reliability judgement,
- the agents' keeping trace of the original information source.

We've developed a simulation testbed based on MICE [16], a specific tool for multi-agent applications in which can be defined various agents with different capabilities, able to communicate each other, perceive the environment, move in it and modify it. Inspiration for our specific testbed has come from CLUE™, a society game based on detective stories metaphor. In the following text we sketch our CLUEDO testbed.

In the nine rooms of the Tudor house move nine characters, each able to use any of the nine arms available in the house. It is common knowledge that the householder is dead. The characters must detect which of their companions is the killer, which weapon he used and in which room he murdered the householder. The clues are randomly placed in the house. Each agent investigates by itself moving around following simple strategies to find out the clues. Any agent has a limited visibility; an agent finds a clue if he can view it. Each agent has a fixed perceptive *capacity*, that is, even if they see a clue, *agents can perceive it badly*. The probability of a wrong perception is proportional to the agent's capacity. On meeting each other, agents exchange their current results. Each agent has a degree of *sincerity*: an agent is sincere if it comunicates the context that it most believes in; an agent is mendacious if it comunicates beliefs it does not believe in. When agents communicate their results they do not specify who found the clues. They act as the clues were personally discovered. Any receiving agent considers the sending agent as the source of the Information. Any agent's database can become inconsistent for two reasons:

1. observing a clue, the agent can perceive it badly,
2. the agent receives incorrect information from mendacious or incapable agents.

On detecting a contradiction in its own database any agent starts a BR process. The agent will:

1. list all the nogoods and goods,[2]
2. calculate the reliability of the other agents and the credibilities of all the clues in its KB,
3. select the more credible good.

If an agent A communicates with B and then changes its preferred good, this change will not affect the preceding communication to B until they will meet again. When an agent's preferred good singles out a solution (one killer, one arm, one room) then the agent proposes it.

### 3.2 Results

Almost seven hundred hours of simulation on a SPARKstation 2 showed that the model doesn't assure convergency when many agents are unreliable. We changed the number of unreliable agents and their

[1] We set ourselves far from [15] which imposes that only facts not depending on communicated information may be sent to other agents.

[2] in the CLUEDO testbed there are no derived sentences, so the concept of "context" collapses to that of "good."

degrees of sincerity and capability, each time running three simulations with different initial dispositions of agents and clues. The following results refer to the case in which all the agents are sincere and only one (agent 2) is incompetent (incapable to perceive correctly the clues):

1) all the agents firmly converge to the same solution of the case;

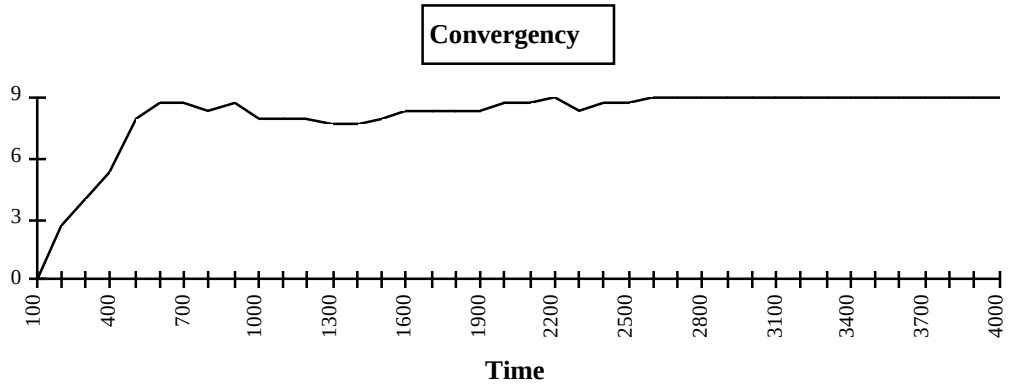


Fig.2 Maximum number of agents proposing the same solution of the case

2) the solution to which they converge is *prevalently* the correct one:

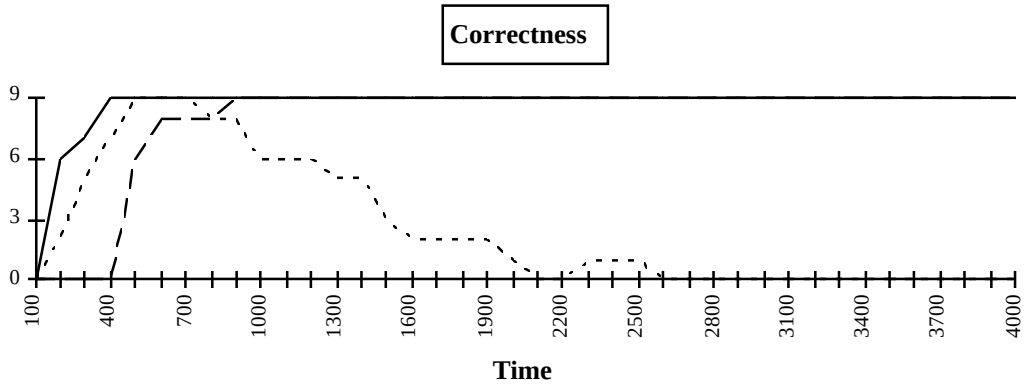


Fig.3 Number of agents proposing the correct solution (in three different classes of simulation)

3) the incompetent agent (agent 2) is averagely recognized as the least reliable one by the others.

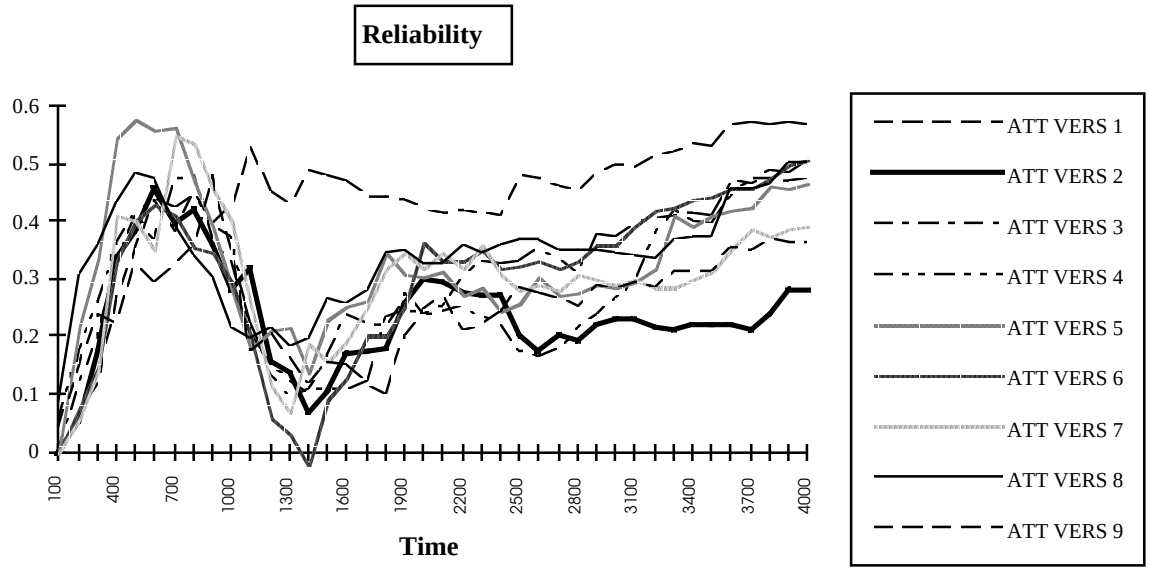


Fig.3 Average of the reliability estimated by eight agents toward the ninth one.

## 4 Conclusions

In this paper we've outlined a distinction between Truth Maintenance (TM) and Belief Revision (BR). BR has here been defined as TM plus some criteria to select the more credible consistent set of beliefs to reason with. In the same way we've distinguished Distributed TM and Distributed BR. Distributed BR has a more complex conceptualization than Distributed TM and it needs the evaluation of special features to behave acceptably from a global perspective. In particular it is important to study how the reliability of a source affects the credibility of the Information and vice-versa. We've proposed a general framework for BR in a multi-agent environment with some parametric functions to relate these features and we've presented a specific multi-agent simulation testbed to study and compare their performances. The results obtained with the simulator confirm that our local BR model favours the opinions convergency, its stability and its correctness without the agents exchanging opinions concerning the others' reliabilities and without keeping trace of the original sources of the information.

## References

[1] Readings in Distributed Artificial Intelligence, A. H. Bond and L. Gasser eds, Morgan Kaufmann Publishers, San Mateo, CA, 1988.

[2] Tom Bylander, D. Allemang, M. C. Tanner and J. R. Josephson, The computational complexity of abduction, *Artificial Intelligence*, 49,25-60, 1991.

[3] Castelfranchi, C., Miceli, M., Cesta, A., Dependence Relations among Autonomuos Agent. In E. Werner & Y. Demazeau (Eds.), *Decentralized A. I. 3*. Elsevier Science Publisher, 1992.

[4] Philip Cohen & Jerry Morgan & Martha Pollack, *Intentions in Communication*, The MIT Press, Cambridge, Mass., 1990.

[5] de Kleer, J., An Assumption Based Truth Maintenance System, *Art.Int.* 28, pp. 127-162, 1986.

[6] de Kleer, J., Focusing on Probable Diagnoses, Proceedings of 1991 Conference of the American Association for Artificial Intelligence, pp 842-848, 1991.

[7] de Kleer, J., Mackworth, A. K., Reiter, R., Characterizing Diagnoses, Proceedings of 1990 Conference of the American Association for Artificial Intelligence, pp 324-330, 1990.

[8] Aldo Franco Dragoni, A Model for Belief Revision in a Multi-Agent Environment. In E. Werner & Y. Demazeau (Eds.), *Decentralized A. I. 3*. NH Elsevier Science Publisher, 1992.

[9] Julia Rose Galliers, Modelling Autonomous Belief Revision in Dialogue, Tech Rep. Cambridge University Comp. Lab., Cambridge (England), 1989.

[10] Joao P. Martins, Stuart C. Shapiro, A Model for Belief Revision, *Artificial Intelligence* 35 (1), pp. 25-79, 1988.

[11] Appelt, D. E. and Pollack, M. E., Weighted Abduction for Plan Ascription, User Modeling and User-Adapted Interaction, vol 2, $n^{OS}$ 1-2, pp. 1-26, Kluwer Academic Publishers, 1992.

[12] Provan, G. M., A Complexity Analysis of Assumption-Based Truth Maintenance Systems, in Smith, B. and Kelleher, G. (Eds.), *Reason Maintenance Systems and Their Applications*, Ellis Horwood Series in Artificial Intelligence, 1988.

[13] Huhns, M. N., Bridgeland, D. M.: Distributed Truth Maintenance. In Dean, S. M., editor, Cooperating

Knowledge Based Systems, pages 133-147. Springer-Verlag, 1990.

[14] Kraetzschmar, G., Beckstein, C., Fuhge, R.: Supporting Assumption-Based Reasoning in a Distributed Environment, in Proceedings of the 12th International Workshop on Distributed Artificial Intelligence, Hidden Valley, Pennsylvania, May 19-21, 1993.

[15] Cindy L. Mason and Rowland R. Johnson, DATMS: A Framework for Distributed Assumption Based Reasoning, in L. Gasser and M. N. Huhns eds., Distributed Artificial Intelligence 2 (Pitman/Morgan Kaufmann, London, pp 293-318, 1989.

[16] T. A. Montgomery, J. Lee, D. J. Musliner, D. E. Damouth, Y. So and E. H. Durfee, *MICE Users Guide*, Artificial Intelligence Laboratory, Dept. of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, February 1992.

[17] R. Reiter "A logic for default reasoning", *Artificial Intelligence*, 13 (1980) 81-132

[18] Reid G. Smith and Randall Davis, Frameworks for Cooperation in Distributed Problem Solving, *IEEE Transactions on Systems, Man and Cybernetics*, SMC-11(1):61-70, 1981.

[19] Doyle, A Truth Maintenance System, *Artificial Intelligence* 12 (3), pp 231-272, 1979.