Modelling Credulity and Skepticism through Plausibility Measures

Berilhes Borges Garcia and Gabriel Pereira Lopes
Departamento de Informática, Universidade Nova de Lisboa

Abstract

In this paper we show how recently observed facts and plausible explanations for them can be accommodated or not in the mental state of an agent. This depends on whether the epistemic surprise degree associated with them is below or above a surprise limit that the agent is willing to accept. This limit is related to the credulity or skepticism degree the agent exhibits. The proposed approach also is sensitive to context known by the agent (its mental state), so that it allows for a given agent to refuse (or accept) a new evidence in a specific context, while it accepts (or refuses) the same belief in a different context. This approach is completely innovative as agents become more flexible in their interaction with the external world.

Keywords: Cognitive Modelling, Nonmonotonic Reasoning, Autonomous Agents.

1 Introduction

The Webster dictionary defines skepticism as being: '1 : an attitude of doubt or a disposition to incredulity either in general or toward a particular object 2 a : the doctrine that true knowledge or knowledge in a particular area is uncertain b : the method of suspended judgement, systematic doubt, or criticism characteristic of skeptics 3 : doubt concerning basic religious principles (as immortality, providence, and revelation)'. Thus skepticism implies unwillingness to believe without conclusive evidence, requiring compelling evidence before we believe. The opposite of the skeptical attitude is called everything from credulity, a relatively kind term, to outright foolishness and self-delusion.

However one can easily observe that the list of behaviours an individual can present cover the whole spectrum from the purest ingenuousness to the most ingrained skepticism. In spite of this apparent consensus, concerning the diversity of an intelligent agent's possible behaviours, practically every research results (Quaresma's [9] exception) on autonomous agents assume an extremely naive posture (credulous); i.e. the agent accepts every belief proposed by the external environment or by any other existing agents, with which it interacts.

This extremely naive posture weakens the agent's rational behaviour. It drives it to non-natural behaviours and into a lot of non-coherent situations. Allowing the modelling the two extreme poles of the spectrum of possible behaviours, i.e. pure ingenuousness and ingrained skepticism, the work of Quaresma and Lopes [10] constitutes a progress in relation the previous approaches. However it is characterized by being highly rigid, leading the agent to behave always in the same manner in every context and does not allow the modelling intermediate behaviours between the two extremes. Thus, in the approach Quaresma and Lopes, a naive agent always accepts to modify its beliefs in function of the new evidence, while a skeptical agent never accepts to modify them.

The main goal of this paper is to propose a new framework which allows to model the whole range of an agent's possible behaviours, and at the same time has into account the context known by the agent. It also allows for a given agent to refuse (or accept) a proposed belief in a specific context, while it accepts (or refuses) the same belief in a different context, possibly due to the evolution of the previous context by adding some new evidence.

In this paper will take as central the notion of plausibility degree presented by Garcia and Lopes [5], and on top of that we will define what does it mean to say that a certain agent is very credulous, or any other gradation with relation to the agent credulity, or skepticism. Intuitively we will say, for example, that an agent is very credulous when it accepts to modify its beliefs in order to accommodate a new observation despite the fact that the most plausible explanation for the observation is very surprising, or in other words it has very little plausibility.

Previous paragraph highlights another basic aspect of our proposal the notion of explanation for the newly observed fact. According to our proposal, when an agent observes a new fact ϕ, it should look for an explanation for this fact, before changing its beliefs. If it is possible to find an explanation α for the newly observed fact, it should additionally determine the plausibility degree for this explanation. If this plausibility degree is above a threshold the agent is willing to accept, it should then change its beliefs. This change is done such way that both ϕ and α are incorporated in agent's belief set as well as the logic consequences of the new fact and of its explanation.

On the other hand, if the agent can not determine an explanation, its beliefs remain unaltered; this means that the agent assumes a cautious position. Of course this can be changed, but the agent could assume an inquisitive posture (similar to ELISA [11]) asking its interlocutor for some kind of explanation. But this will be worked in future work. This scenario is equivalent to the situation where the agent prefers the information in background to the newly observed fact.

This paper is structured in the following way: next section introduces a motivational example and the formalism used to describe our domain. Section 3 presents concisely our proposal for incorporating new evidence in an agents's knowledge base. This proposal is based on the semi-qualitative notion of plausibility and in the abductive explanation for the newly observed fact. Section 4 demonstrates how can the framework previously presented be used for mod-

elling several types of behaviour an agent can show. In the final section we draw the main conclusions and directions for further research.

2 Domain Descriptions

Technically, we start describing the specific knowledge an agent possesses about a certain context. This knowledge T will be represented by a pair (K_D, E), where K_D represents the background knowledge about the domain the agent has, i.e. the generic knowledge about the domain the agent possesses. E represents the contingential knowledge, i.e. the knowledge that is likely to vary from case to case and along the time axis. We also refer to the pair $T = (K_D, E)$ as the known context of the agent or its knowledge base.

Let $\mathcal{L}$ be a set of arbitrary ground literals (an atom a or its explicit negation, $\neg a$). By a schema $\lambda(X)$, where X is a tuple with free variables and sometimes constants, we mean the set of all ground instances $\lambda(a)$ presents in $\mathcal{L}$, where a is a tuple of ground terms, so that every free variable in X is substitued by a ground term.. The background knowledge K_D will be represented by means of a set of default clauses[1]: $\alpha_i \rightsquigarrow \beta_i$. Each default $\alpha_i \rightsquigarrow \beta_i$ is interpreted as a defeasible rule, which can be understood, as stating: "If α_i then normally / typically β_i holds". Where β_i is a schema or its negation. And α_i is a formula constructed from a set of schemas and from the connectives $\vee$, $\wedge$ and $\neg$. $\rightsquigarrow$ is a meta-connective, meaning normally / typically. The additional symbol $\bot$ represents logical falsity.

Example 1 *Take an agent A, having the following specific knowledge, regarding a certain domain:*

$$K_D = \left\{ \begin{array}{l} d_1 : cs(X) \rightsquigarrow \neg int(X, lp) \\ d_2 : cs(X) \rightsquigarrow \neg int(X, lin) \\ d_3 : int(X, ai) \rightsquigarrow int(X, lp) \\ d_4 : int(X, ai) \rightsquigarrow \neg int(X, lin) \\ d_5 : int(X, ai) \rightsquigarrow cs(X) \\ d_6 : int(X, pr\_cl) \rightsquigarrow int(X, lin) \\ d_7 : int(X, pr\_cl) \rightsquigarrow int(X, ai) \end{array} \right\} \quad (1)$$

Rules d_i represent the following facts: (d_1) A believes that computer science (cs) students are normally neither interested in learning logic programming (lp), (d_2) nor linguistics (lin), (d_3) students interested in artificial intelligence (ai) are normally interested on learning logic programming, (d_4) but typically are not interested in learning linguistics, (d_5) students interested in artificial intelligence are normally students from computer science, (d_6) students interested in doing their final course project on computational linguistics (pr_cl) are typically interested on learning linguistics and (d_7) are normally interested in artificial intelligence.

[1] For simplicity reasons and space in this paper we only consider domains described exclusively by default rules.

Assume now that the agent A also knows that b is a computer science student, which can be represented by the contingential knowledge: $E = \{cs(b)\}$.

If after a while A gets evidence in favor of the fact that b is interested in studying linguistics, which is represented by the formula ϕ,

$$\phi = int(b, lin) \tag{2}$$

Should A change its beliefs in order to incorporate this new evidence ϕ? In a general way all the approaches for modelling autonomous agents assume that the answer to this question is positive, regardless of the behaviour characteristics of the agent (that is if the agent is skeptical, not very naive, etc.) and of the context known by the agent. However we conjecture in this work that an agent prefers to maintain his beliefs in background and question the validity of the new evidence, when the most plausible explanation for this new evidence is above the plausibility limit that the agent is willing to accept.

Assumes that an agent A is a credulous agent and that the explanation most plausible for $\phi = int(b, lin)$ is $\alpha = \{int(b, pr\_cl)\}$. This explanation is very surprising (the definition of epistemic surprise and the mechanism to measure it are presented in section 3, however greater details can be found in [5] and [4]) given the context known by agent. In this case we conjecture that the agent's new belief set should incorporate not only the logic consequences of the new fact but also the logic consequences of its explanation. So that the belief set of A is equal to:

$$BS_1 = \{cs(b), int(b, lin), int(b, ai), int(b, pr\_cl)\} \tag{3}$$

Assume now that an agent $A^{'}$ is a bit less credulous than A. It has also the same background knowledge (1) and the same contingential knowledge $E = \{cs(b)\}$ as A. What will be the epistemic state of $A^{'}$ after it observed the new evidence $\phi = int(b, lin)$? Given that the context T known by two agents is identical we are lead to conclude that the more plausible explanation $\alpha = \{int(b, pr\_cl)\}$ to A is also the more plausible explanation to $A^{'}$. Furthermore suppose that the degree of epistemic surprise of $\alpha = \{int(b, ai)\}$ is above of the surprise threshold that $A^{'}$ is willing to accept. Therefore we believe that the belief set of $A^{'}$ is:

$$BS_2 = \left\{ \begin{array}{c} cs(b), \neg int(b, lp), \neg int(b, lin), \neg int(b, ai), \\ goal\left(A^{'}, know\_if\left(int\left(b, pr\_cl\right)\right)\right) \end{array} \right\} \tag{4}$$

Where $goal(A^{'}, know\_if(b, int(b, pr\_cl)))$ represents the fact that $A^{'}$ has the goal of finding out if b is interested in doing his final project course in computational linguistics. In this way the belief set of $A^{'}$ after the observation of ϕ differs from the set of beliefs it possessed before, given that it has now adopted the goal for obtaining additional information in order to clarify the

situation. We can understand this state of beliefs as being an unstable and transitory state, in which the agent sets up a new goal for getting confirmation for the most plausible explanation for ϕ, in order to enable it to incorporate (or reject) new evidence ϕ.

The examples shown above demonstrate that the epistemic state of an agent after the observation of a new evidence ϕ depends on its behavioural tendency, that is if it is very credulous, credulous, etc. However this does not tell the whole story. Another essential aspect is the context known by the agent, that is its background knowledge and its contingential knowledge. The previous example also stresses the central role played by the *most plausible* explanation for a new fact.

3 Plausibility Degree and Abductive Explanations

From the description of the background knowledge K_D of an agent A we use the specificity relations among the default conditionals to establish preference relations between the interpretations that the agent foresees as possible. The determination of the specificity relations is made using system Z [7]; where the set K_D is partitioned into mutually exclusive subsets $K_{D0}, K_{D1}, ..., K_{Dn}$. Two rules belonging to a subset K_{Di} are equally specific. If they belong different subsets they have different specificities. For space reasons we will not detail the partitioning process, thus readers should consult [7], [6] for more details.

Example 2 *The partition of K_D of example (1) is:* $K_{D0} = \{d_1,\ d_2\}$; $K_{D1} = \{d_3, d_4, d_5\}$; $K_{D2} = \{d_6, d_7\}$

The the specificity of the default d_m; written $Z(d_m)$, is i iff $d_m \in K_{Di}$. In the previous example we have $Z(d_1) = Z(d_2) = 0$, $Z(d_3) = Z(d_4) = Z(d_5) = 1$ e $Z(d_6) = Z(d_7) = 2$.

Now, it can be established a relationship between the default ranking function $Z(.)$ and a function $k(.)$ that ascribes to each interpretation w of K_D an ordinal representing the normality of this interpretation. Where an interpretation is a truth assignment to the elements of $\mathcal{L}$ present in $K_D \cup E$.

The link between these two functions is made considering the normality of an interpretation as being equal to the of the higher order default that it violates[2] added with one. Thus, the ranking of an interpretation w in which a student b of computer science, $cs(b)$, is interested on learning logic programming , $int(b, lp)$, is $k(w) = 1$, since the ranking of the higher order default violated by this interpretation is 0 (this interpretation violates the default $cs(X) \rightsquigarrow \neg int(X, lp)$, whose ranking is zero). However it should be stressed that we are interested in measuring how much surprised an agent would be by finding an interpretation that would satisfy ψ given that its mental state already satisfied ϕ. This degree

[2] A default $\alpha \rightsquigarrow \beta$ is verified in an interpretation if both α and β are satisfied, and it is violated in an interpretation if α is satisfied but β is not.

of surprise may be defined as being $k(\psi/\phi)$ [7], and it equals the difference between $k(\psi \wedge \phi)$ and $k(\phi)$, or written otherwise:

$$k(\psi/\phi) = k(\psi \wedge \phi) - k(\phi) \tag{5}$$

It is assumed that the degree of plausibility of a formula ψ since the agent already knows ϕ, $Pl(\psi/\phi)$, is 'inversely' proportional to the degree of surprise of ψ, $k(\psi/\phi)$, i.e. given two propositions α and β we will say that α is at least as plausible as β given the contingential knowledge E, $Pl(\beta/E) \leq Pl(\alpha/E)$, iff $k(\alpha/E) \leq k(\beta/E)$.

From plausibility degree Garcia and Lopes [5] define a ζ-translation of the agent's knowledge base $T = (K_D, E)$ into an extended logic program, with two types of negation: explicit negation and negation by default. This translation plays two roles. In first place it provides a efficient method for computing the logical consequences of the agent's known context. Furthermore, it also allows, on a meta-logic level, to measure the degree de plausibility of a formula. For more details see [5] and [4]. The semantics of this program is given by the Well Founded Semantics eXplicit negation (WFSX) [1].

An extended logic program P is a set of rules of the following kind:

$$L_0 \leftarrow L_1 \wedge ... \wedge L_m \wedge not\ L_{m+1} \wedge ... \wedge not\ L_n \tag{6}$$

Where $0 \leq m \leq n$ and $0 \leq i \leq n$ Each L_i is an objective literal. An objective literal is an atom A or its explicit negation $\neg A$. The symbol *not* represents negation-as-failure and *not* L_i is a default literal. Literals are objective literals or default literals and $\neg\neg A \equiv A$.

Example 3 *For example one considers the background knowledge K_D defined by (1). The ζ-translation of this knowledge is equal to the following program*

P_D:

$$
\begin{aligned}
\neg int(X, lp) &\leftarrow cs(X) \wedge not\ int(X, lp) \wedge not\ ab0. \\
\neg int(X, lin) &\leftarrow ex(X) \wedge not\ int(X, lin) \wedge not\ ab0. \\
ab0 &\leftarrow int(X, ai). \\
int(X, lp) &\leftarrow int(X, ai) \wedge not\ \neg int(X, lp) \wedge not\ \ ab1. \\
\neg int(X, lin) &\leftarrow int(X, ai) \wedge not\ int(X, lin) \wedge not\ ab1. \\
cs(X) &\leftarrow int(X, ai) \wedge not\ \neg cs(X) \wedge not\ ab1. \qquad (7) \\
ab1 &\leftarrow int(X, pr\_cl). \\
int(X, lin) &\leftarrow int(X, pr\_cl) \wedge not\ \neg int(X, lin) \wedge not\ ab2. \\
int(X, ai) &\leftarrow int(X, pr\_cl) \wedge not\ \neg int(X, ai) \wedge not\ ab2. \\
lev0 &\leftarrow not\ \neg lev0 \wedge not\ lev1 \wedge not\ lev2 \wedge not\ lev3. \\
lev1 &\leftarrow cs(X) \wedge not\ \neg int(X, lp) \wedge not\ \neg lev1 \wedge not\ lev2 \wedge not\ lev3. \\
lev1 &\leftarrow cs(X) \wedge not\ \neg int(X, lin) \wedge not\ \neg lev1 \wedge not\ lev2 \wedge not\ lev3. \\
lev2 &\leftarrow int(X, ai) \wedge not\ int(X, lp) \wedge not\ \neg lev2 \wedge not\ lev3. \\
lev2 &\leftarrow int(X, ai) \wedge not\ \neg int(X, lin) \wedge not\ \neg lev2 \wedge not\ lev3. \\
lev2 &\leftarrow int(X, ai) \wedge not\ cs(X) \wedge not\ \neg lev2 \wedge not\ lev3. \\
lev3 &\leftarrow int(X, pr\_cl) \wedge not\ int(X, lin) \wedge not\ \neg lev3 \\
lev3 &\leftarrow int(X, pr\_cl) \wedge not\ int(X, ai) \wedge not\ \neg lev3
\end{aligned}
$$

Where the literal lev_j represents the normality degree, so that the higher the index i, the lower a model's normality will be.

We will refer the Well Founded Model (WFM) of extended logic program P_D obtained by means of ζ-translation of $T = (K_D, E)$ by $WFM_{K_D,E}$. It takes into account the implicit specificity relations in agent's knowledge base and it is a maximally normal model, i.e. minimally surprising, which satisfies E.

The obtained logic program can be enlarged so that to allow the agent determine the possible explanations for the new evidence, according to [3] and [8]. For more details see [5].

Once the agent has determined a possible explanation α for a new fact ϕ, it is able to evaluate the plausibility degree associated to this fact ϕ and its explanation α. All it need to do is to evaluate the difference between the normality of $WFM_{K_D,E}$ and the normality of WFM that results from the assimilation of α and ϕ, $WFM_{K_D,E\cup\alpha\cup\phi}$, in other words $k(WFM_{K_D,E\cup\alpha\cup\phi}) - k(WFM_{K_D,E})$. This can be done in a simple way through the difference that exists between the indexes of the literals lev_i's present in $WFM_{K_D,E\cup\alpha\cup\phi}$ and $WFM_{K_D,E}$.

To summarize so far, we are able to determine the degree of plausibility resulting from the assimilation of a fact ϕ and its explanation α. However, our objective is to determine the degree of plausibility of a new fact ϕ. How can this be done? The simple answer to this question is to assume the most plausible explanation for α, i.e. minimally surprising.

Definition 4 (ϕ's plausibility degree) *Let* $\|T \wedge \phi\| = \{\alpha_1, \alpha_2, ..., \alpha_m\}$ *be a set of explanations for* ϕ *in* T *, thus* $k(\phi/E) = \min_{\alpha_i \in \|T \wedge \phi\|} k(WFM_{K_D, E \cup \alpha_i \cup \phi}) - k(WFM_{K_D, E})$.

4 Modelling Behaviours

In the previous section we briefly presented a framework that allows an agent A to evaluate the plausibility degree of a new information ϕ, given that it already knows a context (mental state) $T = (K_D, E)$. Next step establishes the connection between this plausibility measure and the types of behaviour an agent may exhibit. Initially take two agents, A and A', that possess the same background knowledge K_D. Suppose also that both have the same evidential context E, that is, both know the same context $T = (K_D, E)$.

Assume now that both obtain simultaneously a new evidence ϕ, which is inconsistent with the beliefs presently retained by them. If agent A modifies its beliefs, in order to incorporate the new evidence and its most plausible explanation α, and if A' prefers to engage in an investigation process to guarantee more plausibility for the most plausible explanation α for ϕ. Observe that we can state that the most plausible explanation α found by A will also be the most plausible explanation found by A', given that both agents know the same context T. Taking into account what has been said in section 3 we can conclude that both agents have the same degree of epistemic surprise in relation to the new evidence ϕ.

Being so, how can we justify that both agents show differentiated behaviours? A possible answer, advocated by us in this paper, is that both agents possess different limits of tolerance to epistemic surprise. And that an agent only accepts a new evidence ϕ if the degree of surprise associated to this new evidence is situated below the limit of epistemic surprise that it is willing to accept. Thus, the greater the limit of epistemic surprise an agent has, the higher its capability for accepting a new evidence ϕ will be, that is the more credulous the agent is, and consequently smaller the level of skepticism it will have. The main justification for this assertion is that the observation of ϕ has an informative value that should not be dismissed by the agent.

Proposition 5 (Acceptance of new evidence ϕ) *A new evidence* ϕ *is only accepted by an agent A in a context* $T = (K_D, E)$ *iff* $k(\phi/E)$ *is below the surprise limit it is willing to accept.*

Example 6 *Assume that two agents A e* A' *have the same background knowledge* K_D *defined by (1). Given that the contingential knowledge the agents know is* $E = \{cs(b)\}$ *suppose that the new evidence* $\phi = \{int(b, lin)\}$ *is observed. The abductive framework* P_A *obtained, according to [5], for both agents, in this situation is equal to:*

$$P_A = \left\langle \begin{array}{c} \{P_D \cup cs(b) \leftarrow\}, \{cs(b), int(b, ai), int(b, pr\_cl)\}, \\ \{\bot \leftarrow not\ int(b, lin)\} \end{array} \right\rangle \quad (8)$$

Where P_D is the extended logic program showed in (7), $\{cs(b), int(b,ai), int(b,pr\_cl)\}$ is the abductive literals set and $\{\bot \leftarrow not\ int(b,lin)\}$ denotes the integrity constraints set. In this case the abductive answer is $\{int(b,pr\_cl)\}$. And the plausibility, or better, its epistemic surprise, is respectively $k(int(b,pr\_cl)/E) = 2$, para both agents. Assuming that the surprise threshold of A is 2, we would conclude that the new context known for A is equal to:

$$T_1^A = \{K_D, E \cup \phi \cup int(b,pr\_cl)\} \tag{9}$$

We can see that the conjunction of P_D (the ζ-translation of K_D) with $\{cs(b) \leftarrow\ , int(b,pr\_cl) \leftarrow\ , int(b,lin) \leftarrow\ \}$ derives the following belief set:

$$BS_A = \left\{ \begin{array}{c} cs(b), int(b,lin), int(b,ai), \\ int(b,pr\_cl), ab0, ab1, lev2 \end{array} \right\} \tag{10}$$

which corresponds to advocated intuition. Furthermore if we assume that the surprise threshold of $A^{'}$ is equal to 1, so that in accordance with the proposition (5) the epistemic state of $A^{'}$ does not include the new evidence ϕ. However in this work, we conjecture that agent $A^{'}$ shall acquire a new goal for obtaining an explicit confirmation for the most plausible explanation it has found for the newly observed fact. The main justification for this position is that the observation of ϕ has an informative value that should not be dismissed by the agent. Therefore we believe that the new context known for $A^{'}$ is equal to:

$$T_1^{A^{'}} = \left\{K_D, E \cup goal\left(A^{'}, know\_if\left(int\left(b, pr\_cl\right)\right)\right)\right\} \tag{11}$$

Again we can see that the conjuction of P_D with $\left\{cs(b) \leftarrow\ , goal\left(A^{'}, know\_if\left(int\left(b, pr\_cl\right)\right)\right) \leftarrow\ \right\}$ derives the following belief set:

$$BS_{A^{'}} = \left\{ \begin{array}{c} cs(b), \neg int(b,lin), \neg int(b,lp), \\ goal\left(A^{'}, know\_if\left(int\left(b, pr\_cl\right)\right)\right), lev0 \end{array} \right\} \tag{12}$$

Therefore, we can model the various types of behaviour an agent may present by fixing different values for its limit of epistemic surprise. The only restriction to be imposed refers to the relation of order (rank), so that for any two agents A and $A^{'}$, if A has a higher threshold than $A^{'}$, then A is more credulous than $A^{'}$. In other words, A is more credulous than $A^{'}$ iff $thr\,(A) > thr(A^{'})$, where $thr\,(X)$ represents the limit of epistemic surprise acceptable by the agent X. Another point to be stressed concerns the role played by the context in the process of assimilation of new evidence. Consider again agent $A^{'}$ from the previous example. Assume now that his evidential knowledge is equal to $E^{'} = \{cs(b), int(b,ai)\}$ and it obtains the same evidence $\phi = \{int(b,lin)\}$ as in the previous example. Observe, however, that the plausibility degree of ϕ given $E^{'}$ is equal to 1, $k\left(\phi/E^{'}\right) = 1$. So, according to proposition (5) the new context known by $A^{'}$ is:

$$T_1^{A^{'}} = \left\{K_D, E^{'} \cup \phi \cup int(b,pr\_cl)\right\} \tag{13}$$

In this way, we can conclude that the contingential knowledge E held by an agent before the observation of a new evidence ϕ is a determinant factor in the epistemic state that results from this observation.

5 Conclusion

Along this paper we present a framework that is able to model various types of behaviour an agent may exhibit. This framework relies on two basic notions: the notion of abductive explanation for a fact recently observed and the concept of degree of surprise associated to this new evidence. After, we introduce the notion that a new evidence can only be accepted by the agent, without major questioning, if the degree of surprise of this new evidence is below the limit of epistemic surprise that the agent is willing to accept. Thus, through the assignment of differentiated values to this limit, we can model various types of behaviour.

This framework has two main characteristics. First, it can model various types of behaviour an agent may exhibit. This characteristic, in our opinion, is extremely important when we consider the possibility of a computational agent interacting with other agents, who may convey intentionally or unintentionally erroneous information. In any of these situations, the agent should possess auto-protection mechanisms for its own beliefs. One possible mechanisms is to adopt skeptical positions in relation to its interlocutors. However, we also expect the agent to be sensitive to the context, so that the acceptance, or rejection, of new evidence is conditioned by the context known by the agent (by its mental state). In this way, the agent has a more flexible behaviour, and so it does not necessarily accepts or rejects always a new evidence, but it evaluates the convenience of the new information regarding the context already known. This flexibility is the second main characteristic of our proposition.

Presently we are investigating how this proposition may be applied in the context of advising dialogues, where the process of transmission of beliefs between the interlocutors is mediated by the framework we present here. And when an autonomous agent receives a statement whose degree of plausibility is below its acceptance level, it should try to engage itself in pro-active attitudes in order to clarify the situation. However some questions remain open, namely the ones related to the nature of the surprise limit and how this can be determined.[12]

All the examples that were presented in this paper were experimented with the latest version of the program REVISE [2], a programming system in extended logic for the revision of belief base. This program is based in top-down derivation procedures for WFSX (Well Founded Semantics with eXplicit negation) [1].

References

[1] J. J. Alferes and L. M. Pereira. *Reasoning with Logic Programming.* Springer-Verlag, 1996.

[2] C. V. Damásio, L. M. Pereira, and M. Schroeder. Revise progress report. In *Workshop on Automated Reasoning: Bridging the Gap between Theory and Practice*, University of Sussex, Brighton, 1996.

[3] K. Eshghi and R. A. Kowalski. Abduction compared with negation by failure. In M. Levi, Giorgio; Martelli, editor, *Proceedings of the 6th International Conference on Logic Programming (ICLP '89)*, pages 234–254, Lisbon, Portugal, June 1989. MIT Press.

[4] B. B. Garcia and G. P. Lopes. Incorporating specificity in extended logic programs for belief revision. In D. J. Cook, editor, *Proceedings of the Eleventh International Florida Artificial Intelligence Research Symposium Conference*, pages 215–219, Menlo Park, California,USA, 1998. AAAI Press.

[5] B. B. Garcia and G. P. Lopes. Introducing plausibility measures to the process of belief revision through extended logic programs. In H. Prade, editor, *Proceedings of the 13th European Conference on Artificial Intelligence*, pages 378–382, West Sussex, England, 1998. John Wiley & Sons.

[6] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84(1-2):57–112, 1996.

[7] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In R. Parikh, editor, *TARK: Theoretical Aspects of Reasoning about Knowledge*, pages 121–136. Morgan Kaufmann, 1990.

[8] L. M. Pereira, J. N. Aparício, and J. J. Alferes. Nonmonotonic reasoning with well founded semantics. In K. Furukawa, editor, *Proceedings of the 8th International Conference on Logic Programming*, pages 475–489. MIT, June 1991.

[9] P. Quaresma. *Inference of Attitudes in Dialogue situation.* PhD thesis, Department of Computer Science, Universidade Nova de Lisboa, 1997.

[10] P. Quaresma and J. G. Lopes. Unified logic programming approach to the abduction of plans and intentions in information-seeking dialogues. *Journal of Logic Programming*, 24(1–2):103–119, July/Aug. 1995.

[11] J. Weizenbaum. ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–44, Jan. 1966.

[12] M. Winslett. Reasoning about action using a possible models approach. In *Proceedings of the 7th National Conference on Artificial Intelligence*, pages 89–93. AAAI Press/MIT Press, 1988.