Goal Generation with Relevant and Trusted Beliefs

Célia da Costa Pereira and Andrea G. B. Tettamanzi
Università degli Studi di Milano
Dipartimento di Tecnologie dell'Informazione
Via Bramante 65, I-26013 Crema (CR), Italy
pereira@dti.unimi.it, andrea.tettamanzi@unimi.it

ABSTRACT

A rational agent adopts (or changes) its *goals* when new information (*beliefs*) becomes available or its *desires* (e.g., tasks it is supposed to carry out) change. In conventional approaches to goal generation in which a goal is considered as a "particular" desire, a goal is adopted if and only if *all* conditions leading to its generation are satisfied. It is then supposed that all beliefs are equally *relevant* and their sources completely *trusted*.

However, that is not a realistic setting. In fact, depending on the agent's trust in the source of a piece of information, an agent may decide how strongly it takes into consideration such piece of information in goal generation. On the other hand, not all beliefs are equally relevant to the adoption of a given goal, and a given belief may not be equally relevant to the adoption of different goals.

We propose an approach which takes into account both the relevance of beliefs and the trust degree of the source from which the corresponding piece of information comes, in desire/goal generation. Two algorithms for updating the mental state of an agent in this new setting and three ways for comparing the resulting fuzzy set of desires have been given. Finally, two fundamental postulates any rational goal election function should obey have been stated.

Categories and Subject Descriptors

I.2.3 [**Artificial Intelligence**]: Deduction and Theorem Proving—*Nonmonotonic reasoning and belief revision*

General Terms

Theory

Keywords

Goal generation, beliefs and desires, fuzzy logic

1. INTRODUCTION AND MOTIVATION

Although there has been much discussion on belief change, goal change has not received much attention. Most works on goal *change* found in the literature do not build on results on belief change. That is the case of [2], in which the authors propose a formal representation for goals as rational desires and introduce and formalize dynamic goal hierarchies, but do not formalize explicitly beliefs and plans; or of [16], in which the authors propose an explicit representation of goals suited to conflict resolution based on a preference ordering of sets of goals. Another approach is [15], which models a multi-agent system in which an agent adopts a goal if requested to do so and the new goal is not conflicting with existing goals. This approach is based on goal persistence, i.e., an agent maintains its goals unless explicitly requested to drop them by the originating agent. The main lack of the above approaches is that they suppose that an agent does not use its own mental state for updating goals in a general way. One of the first approaches in this line is that proposed by Thomason [17] whose objective was to describe a formalism designated to integrate reasoning about desires and planning. His motivation was twofold: (i) reflecting on the need to extend planning formalisms to allow inferred goals; and (ii) explaining the need to extend a bare logic of belief and desire to a true system of practical reasoning by adding the capability of reasoning about actions. The work proposed by Broersen and colleagues [3], introduces the BOID architecture in which goals are generated from the conditional mental attitudes beliefs, obligations, intentions and desires. Also the approach proposed by Dignum and colleagues [8], and more recently the one proposed by us [7] are very much in this line. However, (i) the fact that a belief might be more or less relevant to generating a given goal; and (ii) the influence of the reliabilities of the sources of information, are rarely considered.

In this work, we consider the direct relevance relation among beliefs with respect to a given desire/goal, and how this relation influences goal generation. Here, the influence of a belief depends on two factors: the importance of beliefs, and how strongly the agent trusts the source of information. In standard works on goal generation, beliefs were represented by beliefs which must be true (or false) for generating a goal. If one of the beliefs does not abide by these requirements, the relevant goal is not generated. This is a strong restriction. Indeed, in real life, depending on the importance an agent gives to each belief related to a goal, and on the trust it has in the sources of information, it may decide to generate the goal even if not all those conditions are verified.

Let us consider the following example. Suppose you are looking for a house and you cannot go to the place to take a look. Of course, you have some preferences concerning the house you would like to buy. Let us suppose that those

Cite as: Goal Generation with Relevant and Trusted Beliefs, Célia da Costa Pereira and Andrea G. B. Tettamanzi, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril,Portugal,pp.397-404.

preferences are expressed by the following rule: "if the house has a garden, is near the center of the town, and if it is not close to an airport, I would like to buy it". If completely trusted sources tell you that the house has a garden and is near the center but close to an airport, what will you do? If the fact that the house is not close to an airport is the most important requirement for you, you will never buy that house. Instead, if you deem more important to have a house with a garden, it would not be unthinkable that you adopt the desire to buy that house even if it is close to an airport.

Now, let us suppose that information about the house you are interested in comes from different sources you trust to different degrees. Let us suppose that your most important requirement is that the house has a garden. If the first and untrusted source, tells you that the house has a garden, even if the two other and completely trusted sources tell you respectively that the house is near the center and not close to an airport, it is not unthinkable that you do not adopt the desire to buy that house anyway.

In this paper, we attempt to take into account that kind of considerations in goal generation. The relevance among beliefs with respect to a desire are defined as an order relation while we use fuzzy logic [21] to represent degrees of trust.

The paper is organized as follows. Section 2 presents the fuzzy logic-based formalism which will be used throughout the paper. Section 3 illustrates how changes due to the arrival of new information and/or a new desire influences both the agent's sets of beliefs and desires. In Section 4, we propose three methods for comparing the generated sets of desires. In Section 5, the notion of goal set, which is one of the most preferred sets of desires, is defined and two fondamental postulates for a goal set election function are established. Section 6 concludes and discusses the future work.

2. THE FORMALISM

Desires (or motivations) are necessary but not sufficient conditions for action. When a desire is met by other conditions that make it possible for an agent to act, that desire becomes a *goal*. Therefore, given this technical definition of a desire, all goals are desires, but not all desires are goals.

We distinguish two crisp sets of atomic propositions (or atoms): the set $\mathcal{D}$ of all possible desires and the set $\mathcal{K}$ of all possible knowledge items. For the sake of simplicity, we make the assumption that desires and knowledge items are on completely different levels: a desire is not a piece of knowledge and *vice versa*. However, desires can depend on knowledge, while knowledge never depends on desires.

2.1 Basic Considerations

2.1.1 Fuzzy Sets

Fuzzy sets allow the representation of imprecise information. Information is imprecise when the value of the variable to which it refers cannot be completely determined within a given universe of discourse. Fuzzy sets are then a generalization of classical sets obtained by replacing the characteristic function of a set A, χ_A, which takes up values in $\{0,1\}$ ($\chi_A(x) = 1$ iff $x \in A$, $\chi_A(x) = 0$ otherwise) with a *membership function* μ_A, which can take up any value in $[0,1]$. The value $\mu_A(x)$ or, more simply, $A(x)$ is the membership degree of element x in A, i.e., the degree to which x belongs in A.

A fuzzy set is completely defined by its membership function. Therefore, it is useful to define a few terms describing various features of this function, summarized in Figure 1. Given a fuzzy set A, its *core* is the (conventional) set of all elements x such that $A(x) = 1$; its *support*, supp(A), is the set of all x such that $A(x) > 0$. A fuzzy set is *normal* if its core is nonempty. The set of all elements x of A such that $A(x) \geq \alpha$, for a given $\alpha \in (0,1]$, is called the α-cut of A, denoted A_α.

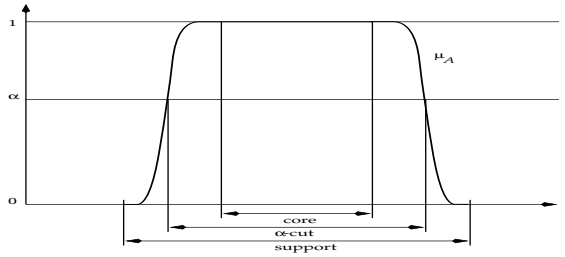

Figure 1: Core, support, and α-cuts of a set A of the real line, having membership function μ_A.

The usual set-theoretic operations of union, intersection, and complement can be defined as a generalization of their counterparts on classical sets by introducing two families of operators, called triangular norms and triangular co-norms. In practice, it is usual to employ the min norm for intersection and the max co-norm for union. Given two fuzzy sets A and B, and an element x,

$$(A \cup B)(x) = \max\{A(x), B(x)\}; \quad (1)$$

$$(A \cap B)(x) = \min\{A(x), B(x)\}; \quad (2)$$

$$\bar{A}(x) = 1 - A(x). \quad (3)$$

2.1.2 Possibility and Necessity Measures

The membership function of a fuzzy set describing imprecise information may be viewed as a *possibility distribution* [22]. Indeed, if A is the fuzzy set of values that a variable x can take up, we denote by $\pi_x = \mu_A$ the possibility distribution attached to x. The identity $\pi_x(u) = \mu_A(u) = A(u)$ means that the degree to which value u belongs to A is the same as the possibility degree of x being equal to u when all is known about x is that it is constrained to take a value in A.

A possibility distribution π induces a *possibility measure* and its dual *necessity measure*, denoted by Π and N respectively. Both measures apply to a crisp set A and are defined as follows:

$$\Pi(A) \equiv \sup_{s \in A} \pi(s); \quad (4)$$

$$N(A) \equiv 1 - \Pi(\bar{A}) = \inf_{s \in \bar{A}} \{1 - \pi(s)\}. \quad (5)$$

In words, the possibility measure of set A corresponds to the greatest of the possibilities associated to its elements; conversely, the necessity measure of A is equivalent to the impossibility of its complement $\bar{A}$.

2.2 Formalism's Components

We present now a fuzzy formalism that accounts for both the relevance and the trust of belief sources. Our formalism

is a fuzzy extension of the formalism proposed in [7] which is an extension of the one proposed in [14]. Like in these approaches, we use a minimal language consisting of atomic beliefs and desires and their negations, and rules expressing relations among beliefs and desires. The extension we are proposing consists of two points:

1. We assume that an agent does not always trust its belief sources completely;
2. We assume that beliefs are differently relevant with respect to the desire to be generated.

Thanks to the first extension proposal, it is possible to represent how strongly the agent believes in a given piece of information. We suppose that this trust degree depends on how reliable is the source of the piece of information[1]. Here, we are not interested in the computation of such reliabilities; we merely assume that, for an agent, a belief has a trust degree in [0, 1]. An approach to the problem of assigning fuzzy trust degrees to information sources can be found for example in previous work by Castelfranchi and colleagues [4, 11].

The second extension proposal, on the other hand, allows us to take into account the fact that a belief could be: (i) crucial for adopting a given desire but less crucial for adopting another different desire; and (ii) more crucial than another in the generation of a particular desire.

Consequently, if we take into account the fact that here the notion of belief is not conceived as an all-or-nothing concept but as a "fuzzy concept", also the relations among beliefs and desires are fuzzy. The fuzzy counterpart of a desire-adoption rule defined in [7] is defined as follows:

Definition 1 (Desire-Adoption Rule). *A desire-adoption rule is an expression of the form* $b_1 \wedge \ldots \wedge b_n \wedge d_1 \wedge \ldots \wedge d_m \rightarrowtail d$, *or* $\delta \rightarrowtail d$, *where* $b_i \in \{p, \neg p\}$ *for some* $p \in \mathcal{K}$, $d_j \in \{q, \neg q\}$ *for some* $q \in \mathcal{D}$, $d \in \mathcal{D}$, $d \neq d_j$ *for all* j, *and* $\delta \in (0, 1]$.

The meaning of the first type of desire-adoption rule which might be called *conditional*, is: "an agent desires d as much as it believes $b_1, \ldots, b_n$ and desires $d_1, \ldots, d_m$. The meaning of the second type of rule, which might be called *unconditional*, is: "the agent (unconditionally) desires d to degree δ".

Given a desire-adoption rule R, we shall denote $\mathrm{lhs}(R)$ the set of literals that make up the conjunction on the left-hand side of R, and $\mathrm{rhs}(R)$ the atom on the right-hand side of R. Furthermore, if S is a set of rules, we define $\mathrm{rhs}(S) = \{\mathrm{rhs}(R) : R \in S\}$.

2.2.1 Relevance and Trust of Beliefs

By taking into account both the *relevance order* and *trust* of beliefs in goal generation it is possible to make trade-offs among beliefs in general, and between most relevant (trusted or untrusted) and less relevant (trusted or untrusted) beliefs in particular, as it happens in real life. Indeed, even if a less relevant (and trusted) belief is false, we often generate a desire anyway or, in the opposite case, even if a highly relevant (but untrusted) belief is true, it would not be unthinkible that we do not generate the desire. It all depends on the relevance and/or trust beliefs have in the process of goal generation.

Let us reconsider the example introduced in Section 1. Suppose you are interested in buying a house, and that your preferences are expressed by the following rule for adopting your desire "to buy a house", *bh*: "if the house has a garden, is near the center of the town, and if it is not close to an airport, I would like to buy it". Let us suppose that

(i) information about the house you are interested in comes from different sources you trust to different degrees, and

(ii) your most important requirement is that the house has a garden.

If a first and completely untrusted source tells you the house has a garden (*hg*), and a second and a third completely trusted sources tell you, respectively, the house is near the center of the town (*hc*) and not close to any airport (*ha*), will you adopt the desire to buy that house?

Following both a conventional approach, in which neither the relevance of the requirements nor the trust of the sources of information are considered, and the approach in which only the influence order of requirements is considered, you would buy that house, because all your preferences are satisfied and you disregard how trustworthy available information is!

Instead, by following the approach we are proposing, in which both the relevance order of requirements and the trust of sources are considered, it is not unthinkable that you would not adopt the desire to buy that house, because the information that your most important requirement is satisfied is not trusted.

In a different scenario, if

(i) your most important requirement is that the house is not close to an airport, and

(ii) the first and the second completely trusted sources tell you, respectively, that the house has a garden and is near the center of the town, but

(iii) a third and completely untrusted source tells you that the house is close to an airport;

will you adopt the desire to buy that house? In this case, it is not unthinkable that you would desire to buy that house because the source of information that the house is close to an airport is untrusted.

We define a trusted belief and the relevance relations among beliefs as follows:

Definition 2 (Trusted Belief). *A trusted belief is a belief* $b \in \mathcal{K}$ *the agent trusts with a degree* $\alpha \in (0, 1]$. *b is completely trusted (untrusted) if* $\alpha = 1$ $(= 0)$.

Here, we suppose that the degree of trust of a belief b corresponds to the degree to which the agent trusts the source of b.

Definition 3 (Relevance Among Beliefs). *Let* $d \in \mathcal{D}$ *be a desire,* $b, b' \in \mathcal{K}$ *be two belief atoms. b is at least as relevant as* b' *for generating desire d, noted* $b \succeq_d b'$, *iff the information brought by b is at least as influencial for generating d as the information brought by* b'. *The relevance order is* strict, *noted* $b \succ_d b'$, *when* $b \succeq_d b'$ *and* $b' \not\succeq_d b$.

[1]Throughout the rest of the paper we make the assumption that the trust of a belief *is* the trust of its source; hence, we will treat the two expressions as synonyms.

In the example, if knowing that the house has a garden were at least as relevant for you as knowing that the house is near the center of the town, this could be represented by stating $hg \succeq_{bh} hc$. Besides, if you could not stand loud noise, this could be represented by $ha \succ_{bh} hg$.

2.2.2 Agent's State

In this section, we first define the mental state of an agent and then answer the following important questions concerning the influence of the mental state in desire adoption:

1. Which beliefs do really matter in desire adoption?
2. When is a desire-adoption rule activated?

The state of an agent is completely described by a triple $\mathcal{S} = \langle \mathcal{B}, \mathcal{R}_J, \mathcal{J} \rangle$, where

- $\mathcal{B}$ is a fuzzy set of atoms (beliefs) on the universe of discourse $\mathcal{K}$;
- $\mathcal{J}$ is a fuzzy set of atoms (desires or motivations) on the universe of discourse $\mathcal{D}$;
- $\mathcal{R}_J$ is a set of desire-adoption rules, such that, for each desire d, $\mathcal{R}_J$ contains at most one rule of the form $\delta \rightarrowtail d$.

The membership degree of a belief atom in $\mathcal{B}$ is the degree to which an agent trusts the information represented by the atom; $\mathcal{B}$ can be naturally extended to literals by noting that $\mathcal{B}(\neg p) = 1 - \mathcal{B}(p)$. $\mathcal{R}_J$ contains the rules which generate desires from beliefs and other more basic desires (subdesires). $\mathcal{J}$ contains all desires which may be deduced from the agents's desire-adoption rule base, given the agent's beliefs and the agent's desires. Besides, we suppose that an agent disposes of a total order $\succeq_d$ on beliefs for every desire d.

We suppose that among the beliefs influencing the process of adopting a desire, the beliefs which really matter are those which both

(i) come from *trusted enough* sources, and

(ii) are *not dominated*.

The justification of hypothesis (i) is that, depending on the desire the agent has to adopt, it might have different threshold degrees to define a source as being trusted [10]. For example, the threshold used by a doctor who has to decide wheather to make a surgical intervention or not, depending on how much she trusts the results of medical analysis of her patient, is different from the threshold she uses if she has to prescribe an antibiotic. The justification of hypothesis (ii) is that beliefs which are not dominated are those which determine the agent's decision to adopt or not a desire. For example, if the belief that the house has a garden, hg, is not dominated, unlike the belief that the house is near an airport, ha, the agent's decision to adopt the desire to buy the house depends on belief hg.

Definition 4 (Trusted Enough Belief). *Let R be a desire-adoption rule, $d = \mathrm{rhs}(R)$ be the desire to be adopted, $b \in \mathrm{lhs}(R)$ be a belief, and α_d be the threshold for adopting desire d: b is said to be trusted enough if and only if $\mathcal{B}(b) \geq \alpha_d$.*

We can extend this definiton to sets of trusted beliefs.

Definition 5 (Trusted Beliefs for a Desire). *Let R be a desire-adoption rule with $d = \mathrm{rhs}(R)$, $\mathcal{B}$ be the fuzzy set of agent's beliefs, and $\mathcal{B}_{\alpha_d}$ be the α_d-cut of $\mathcal{B}$. The current crisp set of trusted enough beliefs for R is*

$$\mathrm{T}_d = \mathrm{lhs}(R) \cap \mathcal{B}_{\alpha_d}.$$

Definition 6 (Non-Dominated Beliefs). *Let R be a desire-adoption rule with $d = \mathrm{rhs}(R)$. The crisp set of non-dominated belief atoms for R is*

$$\mathrm{N}_d = \{b \in \mathrm{lhs}(R) : \neg \exists b' \in \mathrm{lhs}(R) \text{ such that } b' \succ_d b\}$$

From the two above definitions we can now define the beliefs which are really taken into account by the agent.

Definition 7 (Really Relevant Beliefs). *Let R be a desire-adoption rule with $d = \mathrm{rhs}(R)$. The crisp set of really relevant beliefs for R is*

$$\mathrm{RR}_d = \mathrm{T}_d \cap \mathrm{N}_d$$

We can observe in the following propositions, two behaviors very common in real life:

Observation 1. *If all the beliefs in the left-hand side of the rule are both equally trusted and trusted enough, only the non-dominated beliefs matter in the desire-adoption process, i.e., $\mathrm{RR}_d = \mathrm{N}_d$.*

Observation 2. *If all the beliefs in the left-hand side of the rule are non-dominated, only the trusted enough beliefs matter in the desire-adoption process, i.e., $\mathrm{RR}_d = \mathrm{T}_d$.*

We suppose that, when the agent is adopting a desire, it only cares about subdesires and the really relevant beliefs. The other beliefs are not considered. Let us come back to the example about the surgeon who has to make a surgical intervention. She has the following rule: "if the patient is not extremely anemic ($\neg an$), she is not a child ($\neg ch$), she is not allergic to anaesthesia ($\neg al$), and she is in danger of death (dd), the doctor adopts the desire to operate her (op). Let us suppose that the doctor has no subdesires. The desire-adoption rule representing the doctor's reasoning is

$$\neg an \wedge \neg ch \wedge \neg al \wedge dd \rightarrowtail op.$$

Let us suppose that the doctor has the following trust degrees:

- 0.7 to the laboratory which provided the haemochrome exam;
- 1 to herself for judging both if the patient is a child or not, and if the patient is in danger of death;
- 0.8 to the nurse who has filled out the record of patient allergies;

Suppose that the relevance order for deciding whether to make the operation or not is $dd \succeq al \succeq an \succeq ch$, and $al \succeq dd$. Suppose also that:

- the laboratory informs the doctor that the patient is anemic (an);
- the doctor thinks the patient is not a child and she is in danger of death ($\neg ch \wedge dd$); and

- the nurse informs the doctor that the patient is allergic to anaesthesia (al).

If the surgeon deems as being trusted enough only completely trusted information (trust degree equal to 1), by following Definition 7, the really relevant belief is the belief about danger of death, (dd), because both its trust degree is 1, and it is one of the most relevant beliefs. In fact, because $T_{op} = \{ch, dd\}$ and the set of non dominated beliefs is $\mathrm{N}_d = \{dd, al\}$, we have $\mathrm{RR}_d = \{dd\}$. Therefore, the doctor's decision to operate or not depends, in this case, merely on her own belief about the danger of death of the patient; that is, the activation of the corresponding desire-adoption rule depends, in this case, on the single belief dd.

In general, the activation of a rule also depends on its subdesires. For example, in a health system where a doctor is liable for her mistakes, a surgeon must be willing to take on the liability of an operation. Let us represent such willingness by desire o. Then, the desire-adoption rule representing the surgeon's reasoning would become:

$$\neg an \wedge \neg ch \wedge \neg al \wedge dd \wedge o \rightarrowtail op.$$

Therefore, the doctor's decision to operate or not depends not merely on her own belief about the danger of death of the patient but also on her willingness to take on the responsibility of an operation. The activation of the corresponding desire-adoption rule depends then both on belief dd and on the degree to which she is willing to operate, o.

Definition 8 (Degree of Activation of a Rule). *Let R be a desire-adoption rule. The degree af activation of R, $\mathrm{Deg}(R)$, is given by:*

$$\mathrm{Deg}(R) = \min(\min_{b \in \mathrm{RR}_d} \mathcal{B}(b), \min_{d \in \mathrm{lhs}(R)} \mathcal{J}(d)).$$

for conditional desire-adoption rules, and by:

$$\mathrm{Deg}(R) = \delta \text{ if } R = \delta \rightarrowtail d.$$

Remark 1. *In the case of conditional desire-adoption rules,* $\mathrm{RR}_d = \emptyset \Rightarrow \mathrm{Deg}(R) = 0$.

Indeed, if there is no trusted enough belief literal in the left-hand side of a rule, a rational agent does not adopt the corresponding desire at all.

Let us come back to our doctor's example. If the doctor is completely willing to operate (to a degree 1), she absolutely (to a degree 1) adopts the desire to operate her patient. At first sight, that may look extremely risky and not rational; however, when confronted with a situation whereby she believes a patient would die if not operated, a doctor shall operate anyway, even if there is a risk of an anaphylactic shock. This is the only alternative that offers a chance to save the patient's life.

Finally, we do not expect a rational agent to formulate desires out of whim, but based on some rational argument. To model that state of affairs, desire-adoption rules play the role of rational arguments. We define the degree to which a desire is justified as follows.

Definition 9 (Degree of Justification). *The degree of justification of desire d is defined as follows:*

$$\mathcal{J}(d) = \max_{R \in \mathcal{R}_{\mathcal{J}} : \mathrm{rhs}(R) = d} \mathrm{Deg}(R).$$

This degree represents how rational it is the fact that an agent desires d.

3. CHANGES IN THE AGENT'S STATE

The acquisition of a new belief with a given degree of trust in state $\mathcal{S}$ may cause a change in the belief set $\mathcal{B}$ and this may also cause a change in the desire set $\mathcal{J}$ with the variation of the justification degree of desires. Likewise, the arising of a new desire with a given degree may also cause changes in the desire set $\mathcal{J}$.

3.1 Changes Caused by a new Belief

3.1.1 Changes in the Agents's Belief Set

To account for changes in the belief set $\mathcal{B}$ caused by the acquisition of a new belief, we define a new operator for belief change, noted $*$, which is an adaptation of the well known AGM operator for belief revision [1] to the fuzzy belief setting, in the spirit of [19]. The main difference with our work is that we consider literals instead of arbitrary formulas but in addition we also consider the trust degrees of beliefs.

Definition 10 (Belief Change Operator). *Let $*$ be the belief change operator. Let $b \in \mathcal{K}$ be an atomic knowledge item and $\frac{\alpha}{l}$, with $l \in \{b, \neg b\}$, a piece of information concerning b, with a trust degree $\alpha \in [0, 1]$. Let $\mathcal{B}$ be a fuzzy set of beliefs. The new fuzzy set of beliefs $\mathcal{B}' = \mathcal{B} * \frac{\alpha}{l}$ is such that:*

$$\mathcal{B}'(b) = \begin{cases} \mathcal{B}(b) \cdot (1 - \alpha) + \alpha, & \text{if } l = b; \\ \mathcal{B}(b) \cdot (1 - \alpha), & \text{if } l = \neg b. \end{cases} \quad (6)$$

Proposition 1. *If $l = b$, i.e., if the new piece of information does not contradict b, applying the operator $*$ never causes the trust degree of b to decrease, i.e., $\mathcal{B}'(b) \geq \mathcal{B}(b)$.*

Proof. If $\mathcal{B}(b) = 0$ the result is obvious. Otherwise, if $\mathcal{B}(b) > 0$, we have $\mathcal{B}'(b) - \mathcal{B}(b) = \alpha \cdot (1 - \mathcal{B}(b)) \geq 0$. □

Proposition 2. *If $l = \neg b$, i.e., if the new piece of information contradicts b, applying the operator $*$ never causes the trust degree of b to increase, i.e., $\mathcal{B}'(b) \leq \mathcal{B}(b)$.*

Proof. $\mathcal{B}'(b) - \mathcal{B}(b) = -\alpha \cdot \mathcal{B}(b) \leq 0$. □

The semantics of our belief change operator is defined by the following properties. Here $\mathcal{B}$ represents a fuzzy belief set, l an acquired trusted information, supp is the support of a fuzzy set, and $\cup$, $\cap$, $\subseteq$ and $\supseteq$ are fuzzy operators.

- $(\mathbf{P} * \mathbf{1})$(Stability) The result of applying * in $\mathcal{B}$ with l is always a fuzzy set of beliefs:

$$\mathcal{B} * \frac{\alpha}{l} \text{ is a fuzzy set of beliefs.}$$

- $(\mathbf{P} * \mathbf{2})$(Expansion) If $l = b$, the fuzzy set of information expands:

$$\mathrm{supp}(\mathcal{B} * \frac{\alpha}{l}) \supseteq \mathrm{supp}(\mathcal{B}).$$

- $(\mathbf{P} * \mathbf{3})$(Shrinkage) If $l = \neg b$, the fuzzy set of beliefs shrinks:

$$\mathrm{supp}(\mathcal{B} * \frac{\alpha}{l}) \subseteq \mathrm{supp}(\mathcal{B}).$$

- $(\mathbf{P} * \mathbf{4})$(Invariance) If the new information is completely untrusted, i.e., $\alpha = 0$, invariance holds:

$$(\alpha = 0) \Rightarrow (\mathcal{B} * \frac{\alpha}{l} = \mathcal{B}).$$

- $(\mathbf{P}*\mathbf{5})$(Predictability) The result of applying $*$ contains all beliefs in $\text{supp}(\mathcal{B} \cup \{\frac{\alpha}{l}\})$:

$$\text{supp}(\mathcal{B} * \frac{\alpha}{l}) \supseteq \text{supp}(\mathcal{B} \cup \{\frac{\alpha}{l}\}).$$

- $(\mathbf{P}*\mathbf{6})$(Identity) The result of applying $*$ does not depend on the particular information. If $l_1 (\in \{b_1, \neg b_1\}) = l_2 (\in \{b_2, \neg b_2\})$ and $\alpha_1 = \alpha_2$:

$$\mathcal{B} * \frac{\alpha_1}{l_1} = \mathcal{B} * \frac{\alpha_2}{l_2}.$$

3.1.2 Changes in the Agents's Desire Set

The acquisition of a new belief may induce changes in the justification degree of some desires. More generally, the acquisition of a new belief may induce changes in the belief set of an agent which, in turn, may induce changes in its desire set. Let $\frac{\alpha}{l}$, with $l \in \{b, \neg b\}$, be a new belief trusted to degree α. To account for the changes in the desire set caused by this new acquisition, we must consider

(i) each rule $R \in \mathcal{R}_\mathcal{J}$ such that $b \in \text{lhs}(R)$ or $\neg b \in \text{lhs}(R)$, and

(ii) the relevance order of beliefs $b \in \text{lhs}(R)$ with respect to desire $\text{rhs}(R)$, in order to update the justification value of $\text{rhs}(R)$.

The new desire set $\mathcal{J}'$ is obtained by executing the algorithm in Figure 2 with the following inputs: $\mathcal{B}' = \mathcal{B} * \frac{\alpha}{l}$, $\mathcal{R}_\mathcal{J}$, $\mathcal{D}$, and the agent's relevance order on beliefs. The algorithm propagates changes until a fixpoint is reached; C_k is the set of desires whose justification degree changes in step k, i.e., $\forall d \in \mathcal{D}$, $d \in C_k \Rightarrow \mathcal{J}'_k(d) \neq \mathcal{J}'_{k-1}(d)$.

Of course, the set $\mathcal{R}_\mathcal{J}$ does not change.

PROPOSITION 3. *If the new information is a positive literal,*

$$\mathcal{J}' = \bigcup_{k=0}^{\infty} \mathcal{J}'_k.$$

PROOF. According to Proposition 1, for all b we have $\mathcal{B}'(b) \geq \mathcal{B}(b)$. Therefore, the degree of all desires d in the new desire set $\mathcal{J}'$ may not decrease, i.e., for all k, $\mathcal{J}'_k(d) \geq \mathcal{J}'_{k-1}(d)$. □

PROPOSITION 4. *If the new information is a negative literal,*

$$\mathcal{J}' = \bigcap_{k=0}^{\infty} \mathcal{J}'_k.$$

PROOF. According to Proposition 2, for all b we have $\mathcal{B}'(b) \leq \mathcal{B}(b)$. Therefore, the degree of all desires d in the new desire set $\mathcal{J}'$ may not increase, i.e., for all k, $\mathcal{J}'_k(d) \leq \mathcal{J}'_{k-1}(d)$. □

3.2 Changes Caused by a New Desire

The acquisition of a new desire may cause changes in the fuzzy desire set and in the desire-adoption rule base. In this work, for the sake of simplicity, we consider only new desires which are not dependent on beliefs and/or other desires. A new desire, justified with degree δ, implies the addition of the desire-generation rule $\delta \rightarrowtail d$ into $\mathcal{R}_J$, resulting in the new base $\mathcal{R}'_J$. By definition of a desire-adoption rule base, $\mathcal{R}'_J$ must not contain another $\delta' \rightarrowtail d$ with $\delta \neq \delta'$. How does $\mathcal{S}$ change with the arising of the new desire $\frac{\delta}{d}$?

1. $\mathcal{B}' \leftarrow \mathcal{B} * \frac{\alpha}{l}$; $k \leftarrow 1$; $C_0 \leftarrow \emptyset$;
2. For each $d \in \mathcal{D}$ do
   (a) consider all $R_i \in \mathcal{R}_\mathcal{J}$ such that $\text{rhs}(R) = d$;
   (b) calculate $\text{Deg}(R_i)$ by considering $\mathcal{B}'$ and $\succeq_d$;
   (c) $\mathcal{J}'_0(d) \leftarrow \max_{R_i} \text{Deg}(R_i)$;
   (d) if $\mathcal{J}'_0(d) \neq \mathcal{J}(d)$ then $C_0 \leftarrow C_0 \cup \{d\}$.
3. repeat
   (a) $C_k \leftarrow \emptyset$;
   (b) for each $d \in C_{k-1}$ do
      i. for all $R_j \in \mathcal{R}_\mathcal{J}$ such that $d \in \text{lhs}(R_j)$ do
         A. calculate $\text{Deg}(R_j)$ considering $\mathcal{J}'_{k-1}(d)$ and $\succeq_{\text{rhs}(R_j)}$;
         B. $\mathcal{J}'_k(\text{rhs}(R_j)) \leftarrow \max_{R_i | \text{rhs}(R_i) = \text{rhs}(R_j)} \text{Deg}(R_i)$;
         C. if $\mathcal{J}'_k(\text{rhs}(R_j)) \neq \mathcal{J}'_{k-1}(\text{rhs}(R_j))$ then $C_k \leftarrow C_k \cup \{\text{rhs}(R_j)\}$.
      ii. $k \leftarrow k + 1$.
4. until $C_{k-1} = \emptyset$.
5. for all d, $\mathcal{J}'(d)$ is given by the following equation:

$$\mathcal{J}'(d) = \begin{cases} \mathcal{J}(d), & \text{if } d \notin C; \\ \mathcal{J}'_i(d), & \text{otherwise,} \end{cases} \quad (7)$$

where i is such that $d \in C_i$ and $\forall j \neq i$ if $d \in C_j$ then $j \leq i$, i.e., the justification degree of a "changed" desire is the last degree it takes, and $C = \bigcup_{k=0}^{\infty} C_k$ is the set of "changed" desires.

Figure 2: An algorithm to compute the new desire set upon arrival of a new belief.

- Any rule $\delta' \rightarrowtail d$ with $\delta \neq \delta'$ is retracted from $\mathcal{R}_J$,
- $\delta \rightarrowtail d$ is added to $\mathcal{R}_J$,

It is clear that the arising of a new desire does not change the belief set of the agent.

The new fuzzy set of desires, $\mathcal{J}'$, is computed by the algorithm in Figure 3.

4. COMPARING SETS OF DESIRES

As presented in Section 2.2.2, the fuzzy set of current desires $\mathcal{J}$ is one of the components of the agent's mental state. Such a fuzzy set expresses the fact that an agent may have many differently justified desires at the same time. However, it is also essential to represent the fact that not all desires have the same importance or urgency for the agent. This can be naturally represented by the notion of *utility*. We propose two kinds of represetations for desire utilities which depend whether the agent disposes of numerical or ordinal utilities. The first one is inspired by the von Neumann-Morgenstern utility function $u : \mathcal{D} \rightarrow \mathbb{R}$ and associates a real value to all desires [13]. The second one is inspired by the counterpart of the von Newmann and Morgenstern expected utility theory in the framework of possibility theory proposed by Dubois and Prade [9]. In this case, utilities represent a preference order among possibility distributions on desires.

We have to extend those notions to the case of fuzzy desire sets. We distinguish three cases. In the first case we dispose of qualitative utilities; in the second case we dispose of quantitative utilities; in the last case we do not dispose

1. if $\{\delta' \rightarrowtail d\} \in \mathcal{R}_{\mathcal{J}}$ then $\mathcal{R}'_{\mathcal{J}} \leftarrow (\mathcal{R}_{\mathcal{J}} \setminus \{\delta' \rightarrowtail d\}) \cup \{\delta \rightarrowtail d\}$; else $\mathcal{R}'_{\mathcal{J}} \leftarrow \mathcal{R}_{\mathcal{J}} \cup \{\delta \rightarrowtail d\}$;
2. $k \leftarrow 1$; $C_0 \leftarrow \{d\}$; $\mathcal{J}'_0(d) \leftarrow \delta$;
3. repeat
   (a) $C_k \leftarrow \emptyset$;
   (b) for each $d \in C_{k-1}$ do
      i. for all $R_j \in \mathcal{R}'_{\mathcal{J}}$ such that $d \in \text{lhs}(R_j)$ do
         A. calculate their respective degrees $\text{Deg}(R_j)$ considering $\mathcal{J}'_{k-1}(d)$;
         B. $\mathcal{J}'_k(\text{rhs}(R_j)) \leftarrow \max_{R_i | \text{rhs}(R_i)=\text{rhs}(R_j)} \text{Deg}(R_i)$;
         C. if $\mathcal{J}'_k(\text{rhs}(R_j)) \neq \mathcal{J}'_{k-1}(\text{rhs}(R_j))$ then $C_k \leftarrow C_k \cup \{\text{rhs}(R_j)\}$.
      ii. $k \leftarrow k + 1$.
4. until $C_{k-1} = \emptyset$.
5. for all d, $\mathcal{J}'(d)$ is given by Equation 7.

Figure 3: An algorithm to compute the new desire set upon the arisal of a new desire.

of utilites at all. In all three cases, the preference relation between two fuzzy sets of desires is noted $\succeq$.

4.1 Preference under Qualitative Utility

We adapt the notion of pessimistic utilities [9] and optimistic utilities [20] for the purposes of our work.

Definition 11 (Pessimistic Utility). *Let $\mathcal{J}$ be a fuzzy set of desires, and $u : \mathcal{D} \rightarrow [0, 1]$ the function which maps a desire to a qualitative utility. The pessimistic utility of $\mathcal{J}$, $U^{\text{Pes}}(\mathcal{J})$, is given by:*

$$U^{\text{Pes}}(\mathcal{J}) = \min_d \max(1 - \mathcal{J}(d), u(d)).$$

$U^{\text{Pes}}(\mathcal{J})$ is the inclusion degree of the set of justified desires in the set of useful desires. If all desires d in $\mathcal{J}$ are completetly justified, i.e., $\mathcal{J}(d) = 1\ \forall d$, the pessimistic utility of $\mathcal{J}$ equals the utility of the least useful desire.

Definition 12 (Optimistic Utility). *Let $\mathcal{J}$ be a fuzzy set of desires, and $u : \mathcal{D} \rightarrow [0, 1]$ the function wich maps a desire to a qualitative utility. The optimistic utility of $\mathcal{J}$, $U^{\text{Opt}}(\mathcal{J})$, is given by:*

$$U^{\text{Opt}}(\mathcal{J}) = \max_d \min(\mathcal{J}(d), u(d)).$$

$U^{\text{Opt}}(\mathcal{J})$ is the intersection degree of the set of justified desires with the set of useful desires. If all desires d in $\mathcal{J}$ are completetly justified, the optimistic utility of $\mathcal{J}$ equals the utility of the most useful desire.

Definition 13 (Preference). *Given two fuzzy sets of desires $\mathcal{J}_1$ and $\mathcal{J}_2$, $\mathcal{J}_1$ is preferred to $\mathcal{J}_2$, in symbols $\mathcal{J}_1 \succeq \mathcal{J}_2$, iff $U^{\text{Pes}}(\mathcal{J}_1) > U^{\text{Pes}}(\mathcal{J}_2)$; or $U^{\text{Pes}}(\mathcal{J}_1) = U^{\text{Pes}}(\mathcal{J}_2)$ and $U^{\text{Opt}}(\mathcal{J}_1) \geq U^{\text{Opt}}(\mathcal{J}_2)$.*

4.2 Preference under Quantitative Utility

In case we dispose of quantitative utilities, the utility of a fuzzy set of desires may be calculated as follows.

Definition 14 (Quantitative Utility). *Let $\mathcal{J}$ be a fuzzy set of desires, and $u : \mathcal{D} \rightarrow \mathbb{R}$ a function wich maps a desire to a real value. The utility of $\mathcal{J}$ is*

$$U(\mathcal{J}) = \sum_d u(d) \cdot \mathcal{J}(d).$$

$U(\mathcal{J})$ is the average of utilities weighted by degree of membership.

Definition 15 (Preference). *A fuzzy set of desires $\mathcal{J}_1$ is prefered to $\mathcal{J}_2$, in symbols $\mathcal{J}_1 \succeq \mathcal{J}_2$, iff $U(\mathcal{J}_1) \geq U(\mathcal{J}_2)$.*

4.3 Preference without Utilities

In case we do not dispose of utilities, we can still compare sets of desires by using the justification degrees of their elements. We consider two parameters: the possibility $\Pi(\mathcal{J})$ and the necessity $N(\mathcal{J})$ that the fuzzy set $\mathcal{J}$ is justified.

$\Pi(\mathcal{J}) = \max_d \mathcal{J}(D)$ represents how possibly justified is the fuzzy set $\mathcal{J}$. $\Pi(\mathcal{J}) = 0$ means that $\mathcal{J}$ is certainly not a desire set of the agent. $\Pi(\mathcal{J}) = 1$ means that it would not be surprising at all if $\mathcal{J}$ were the desire set of the agent.

$N(\mathcal{J}) = 1 - \max_{d \in \text{rhs}(\mathcal{R}_{\mathcal{J}})}(1 - \mathcal{J}(d))$ represents how surely justified is the set $\mathcal{J}$. That is because we consider only desires which are justified, i.e., desires in the right hand side of a desire-generation rule with a nonzero justification degree, instead of the entire set of possible desires. $N(\mathcal{J}) = 0$ means that it would not be surprising at all if $\mathcal{J}$ were not a set of desires of the agent. $N(\mathcal{J}) = 1$ means that it is certainly true that $\mathcal{J}$ is a set of desires of the agent.

Definition 16 (Preference). *A fuzzy set of desire $\mathcal{J}_1$ is preferred to a fuzzy set of desires $\mathcal{J}_2$, in symbols $\mathcal{J}_1 \succeq \mathcal{J}_2$, iff $N(\mathcal{J}_1) > N(\mathcal{J}_2)$; or $N(\mathcal{J}_1) = N(\mathcal{J}_2)$ and $\Pi(\mathcal{J}_1) \geq \Pi(\mathcal{J}_2)$.*

5. GOAL SETS

Goals serve a dual role in the deliberation process, capturing aspects of both *intentions* and *desires*. Besides expressing desirability, when an agent adopts a goal, it also makes a commitment to pursue the goal. Here, we concentrate exclusively on the second role served by a goal. For more information about intentions see for example Cohen and Levesque [5].

The main point about desires is that we expect a rational agent to try and manipulate its surrounding environment to fulfill them. In general, considering a planning problem $\mathcal{P}$ to solve, not all desires can be fulfilled at the same time, especially when there is no solution plan which allows to reach all of them at the same time.

We assume we dispose of a $\mathcal{P}$-dependent function $\mathcal{F}_{\mathcal{P}}$ wich, given a fuzzy set of beliefs $\mathcal{B}$ and a fuzzy set of desires $\mathcal{J}$, returns a degree γ which corresponds to the certainty degree of the most certain solution plan found [6]. We may call γ the *degree of feasibility* of $\mathcal{J}$ given $\mathcal{B}$, i.e., $\mathcal{F}_{\mathcal{P}}(\mathcal{B}, \mathcal{J}) = \gamma$. In general, a rational agent will try to reach a set of desires which, first of all, has a suitable degree of feasibility. The preference criterion comes into play in a second time.

Definition 17 (γ-Goal Set). *A γ-goal set, with $\gamma \in [0, 1]$, in state $\mathcal{S}$ is a fuzzy set of desires $\mathcal{G}$ such that:*

1. *$\mathcal{G} \subseteq \mathcal{J}$, i.e., for all $d \in \mathcal{D}$, $\mathcal{G}(d) \leq \mathcal{J}(d)$;*
2. *$\mathcal{F}_{\mathcal{P}}(\mathcal{B}, \mathcal{G}) \geq \gamma$.*

Postulates of a γ-Goal Set Election Function

In general, given a fuzzy set of desires $\mathcal{J}$, there may be more than one possible γ-goal sets $\mathcal{G}$. However, a rational agent in state $\mathcal{S} = \langle \mathcal{B}, \mathcal{J}, \mathcal{R}_J \rangle$ will elect as the set of goals it is pursuing one precise goal set $\mathcal{G}^*$, which depends on $\mathcal{S}$.

Let us call G_γ the function which maps a state $\mathcal{S}$ into the γ-goal set elected by a rational agent in state $\mathcal{S}$: $\mathcal{G}^* = G_\gamma(\mathcal{S})$. A goal election function G_γ must obey two fundamental postulates:

- **(G1)** $\forall \mathcal{S}, G_\gamma(\mathcal{S})$ is a γ-goal set;
- **(G2)** $\forall \mathcal{S}$, if $\mathcal{G}$ is a γ-goal set, then $G_\gamma(\mathcal{S}) \succeq \mathcal{G}$.

Postulate **(G1)** requires, as it is obvious, that a goal election function G_γ does indeed return a γ-goal set.

Postulate **(G2)** requires that the γ-goal set returned by function G_γ be "optimal", i.e., that a rational agent always selects one of the most preferrable γ-goal set.

In general, different goal election functions can be defined that respect the above postulates. While the definition of a specific goal election function is a critical part of constructing a rational agent framework, this issue falls out of the scope of this work.

6. CONCLUSION AND FUTURE WORK

Thomason observes that the relation between goals and desires becomes crucial when the goal generation process is considered [17]. Conventional approaches consider goals as a primitive concept which does not derive from other mental atitudes. This hypothesis avoids the possibility to take into account the fact that goals can come from desires; and that desires can be impracticable or mutually conficting.

Being aware of that, we have proposed a new framework for generating goals dealing with beliefs and desires in rational agents. The originality of such formalism is that it allows to take into account two considerations frequently made in real life — relevance of beliefs to a goal and the agent's trust in a belief. We have proposed two algorithms for updating the mental state of an agent in this new setting and three ways for comparing the resulting fuzzy set of desires have been given. Finally, two fundamental postulates any rational goal election function should obey have been stated.

The next step is to extend our language into a full propositional one like those used in BDI approaches, for example [18]. A further step is to define a mapping from BDI mental states to propositional planning in the spirit of the work by Meneguzzi and colleagues [12] in order to construct a sort of formalism combining planning with general-purpose nonmonotonic reasoning about beliefs and desires.

7. REFERENCES

[1] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *J. Symb. Log.*, 50(2):510–530, 1985.

[2] J. Bell and Z. Huang. Dynamic goal hierarchies. In *PRICAI '96: Proceedings from the Workshop on Intelligent Agent Systems, Theoretical and Practical Issues*, pages 88–103, London, UK, 1997. Springer-Verlag.

[3] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly Journal*, 2(3–4):428–447, 2002.

[4] C. Castelfranchi, R. Falcone, and G. Pezzulo. Trust in information sources as a source for trust: a fuzzy approach. In *Proceedings of AAMAS'03*, pages 89–96, 2003.

[5] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artif. Intell.*, 42(2-3):213–261, 1990.

[6] C. da Costa Pereira, F. Garcia, J. Lang, and R. Martin-Clouaire. Planning with graded nondeterministic actions: a possibilistic approach. *International Journal of Intelligent Systems*, 12:935–962, 1997.

[7] C. da Costa Pereira and A. Tettamanzi. Towards a framework for goal revision. In *Proceedings of BNAIC'06*, pages 99–106, 2006.

[8] F. Dignum, D. N. Kinny, and E. A. Sonenberg. From desires, obligations and norms to goals. *Cognitive Science Quarterly .*, 2(3-4):407–427, 2002.

[9] D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. In *Proceedings of IJCAI'95*, pages 19–25, 1995.

[10] R. Falcone and C. Castelfranchi. Social trust: A cognitive approach. In C. Castelfranchi and Y.-H. Tan, editors, *Trust and Deception in Virtual Societibes*, pages 55–90. Kluwer Academic Publishers, pp 55-90, 2001.

[11] R. Falcone and C. Castelfranchi. Trust dynamics: How trust is influenced by direct experiences and by trust itself. In *Proceedings of AAMAS '04*, pages 740–747, 2004.

[12] F. R. Meneguzzi, A. F. Zorzo, and M. da Costa Móra. Propositional planning in BDI agents. In *Proceedings of SAC'04*, pages 58–63, 2004.

[13] J. V. Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[14] I. Rahwan and L. Amgoud. An argumentation-based approach for practical reasoning. In *Proceedings of AAMAS'06*, pages 347–354, 2006.

[15] S. Shapiro, Y. Lespérance, and H. J. Levesque. Goal change. In *Proceedings of IJCAI'05*, pages 582–588, 2005.

[16] J. Thangarajah, L. Padgham, and J. Harland. Representation and reasoning for goals in BDI agents. In *Proceedings of CRPITS'02*, pages 259–265, 2002.

[17] R. H. Thomason. Desires and defaults: A framework for planning with inferred goals. In *Proceedings of KR'00*, pages 702–713, 2000.

[18] M. B. van Riemsdijk. *Cognitive Agent Programming: A Semantic Approach*. PhD thesis, University of Utrecht, 2006.

[19] R. Witte. Fuzzy belief revision. In *Proceedings of NMR'02*, pages 311–320, 2002.

[20] R. Yager. An approach to ordinal decision making. *International Journal of Approximate Reasoning*, 12:237–261, 1995.

[21] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[22] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.