

Goal Generation and Adoption from Partially Trusted Beliefs

Célia da Costa Pereira and Andrea G. B. Tettamanzi¹

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 this paper we propose a non-conventional approach for adopting goals which takes the degree of trust in the sources of information into account. Beliefs, desires, and goals, as a consequence, are gradual. Incoming information may be any propositional formula.

Two algorithms for updating the mental state of an agent in this new setting are proposed. The first algorithm is relevant to the updating when a new piece of information arrives and the second one is relevant to the updating when a new desire arises.

1 Introduction and Motivation

Changes in the mental attitudes of a BDI agent [14] may influence its behavior when deciding which information to believe, which desires/goals to generate/adopt, which action to perform, and so on. The goals to be adopted in a given situation may depend on the agent's beliefs, desires and obligations. However, most works on goal change and generation do not build on results on belief change, e.g., [2, 16, 15]. One of the first approaches in this line is Thomason's [17], whose objective is to describe a formalism designed to integrate reasoning about desires and planning. The work by Broersen and colleagues [3] introduces the BOID architecture in which goals are generated from conditional beliefs, obligations, intentions, and desires. Also the approach by Dignum and colleagues [9] and, more recently, the one proposed in [6] are very much in this line. However, these works consider the notion of belief as an all-or-nothing concept: either the agent believes something, or it does not. Parsons and Giorgini [13] proposed to treat beliefs as degrees of evidence. Hansson [11] pointed out that there are two notions of degree of belief. The first is the static concept of degree of confidence. In this sense, the higher an agent's degree of belief in a sentence, the more confidently it entertains that belief. The other notion is the dynamic concept of degree of resistance to change. In that sense, the higher an agent's degree of belief in a sentence, the more difficult it is to change that belief.

In this work, we just consider beliefs, desires, and goals, not intentions, but, in addition, we allow an agent to believe a piece of information to a degree. Such degree depends on the trust the agent has in the source of information. That way, we make it possible to represent the fact that if a piece of information comes from a completely trusted source, like in traditional approaches, the degree of resistance to change of the agent is null and, therefore, the agent revises its beliefs and (completely) adopts the new belief. Instead, if the

agent does not trust the source at all, its belief will not change. The interesting case is when information comes from a partially trusted source. In this case, we will show that the relative shift of an agent's belief degrees depends only on the degree of trust of the source. Our aim is not to compute such trust degrees, we are just interested in how they influence the agent's beliefs and, as a consequence, the choice of which set of goals, among the possible ones, it will adopt.

This work is an extension of one of the first attempts to study the impact of trusted beliefs on desires and goals [7]. The extension consists in allowing all kinds of information including disjunctive information.

To explain the kind of issues we want to address, let us consider the following example, which we will refer to in the rest of the paper. You go for dinner to a new restaurant. You like to have meat (*hm*) but, if you find fresh fish (*ff*), you'd rather have fish (*hf*). Also, you like to have red wine (*rw*) with meat and white wine (*ww*) with fish. When you go to a new place, you assume fish is not fresh, unless you find evidence to contrary — however, you leave some room to doubt. Now, your friend, who already knows the place and whom you trust pretty much, albeit not completely, tells you they usually have fresh fish or, when they don't, their escargots are great (*ge*), which you would be curious to try (*he*) in case you decided not to have fish.

In this paper, we attempt to take into account this kind of considerations on beliefs and desires in goal generation/adoption.

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 influence the agent's beliefs and desires. In Section 4, the notion of goal set is defined and requirements for goal set adoption are underlined. Section 5 concludes.

2 The Formalism

The formalism which will be used throughout the paper is inspired by the one used in [18]. However, unlike [18], the objective of our formalism is to analyze, not to develop, agent systems. Precisely, our agent must single out an *optimal* set of goals to be adopted.

2.1 Basic Considerations

Fuzzy sets, introduced by Zadeh [19], are a generalization of classical sets obtained by replacing the characteristic function of a set A 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 . The *support* of A , $\text{supp}(A)$, is the set of all x such that $A(x) > 0$.

The usual set-theoretic operations of union, intersection, and complement can be defined as a generalization of their counterparts

¹ Università degli Studi di Milano, Italy, email: {pereira,tettamanzi}@dti.unimi.it

on classical sets by introducing two families of operators, called triangular norms and 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 , $(A \cup B)(x) = \max\{A(x), B(x)\}$, $(A \cap B)(x) = \min\{A(x), B(x)\}$, and $\bar{A}(x) = 1 - A(x)$.

Definition 1 (Fuzzy Interpretation) *A fuzzy interpretation is an assignment of truth degrees in $[0, 1]$ to all atomic propositions (or atoms, for short) defined in the problem domain. Given a set of atoms \mathcal{A} , a fuzzy interpretation is a function $\mathcal{I} : \mathcal{A} \rightarrow [0, 1]$, which assigns a truth degree $\mathcal{I}(p) \in [0, 1]$ to all atoms $p \in \mathcal{A}$.*

Note that a fuzzy interpretation is, in all respects, a fuzzy set of atoms.

2.2 Formalism's Components

An agent's belief is a piece of information that the agent believes in. An agent's desire is something (not always material) that the agent would like to possess or perform.

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. The main distinction we made here between desires and goals is in line with that made by Thomason [17] and other authors: goals are required to be consistent whereas desires need not be.

Definition 2 (Language) *Let \mathcal{A} be a set of atomic propositions and let \mathcal{L} be the propositional language such that $\mathcal{A} \cup \{\top, \perp\} \subseteq \mathcal{L}$, and, $\forall \phi, \psi \in \mathcal{L}$, $\neg\phi \in \mathcal{L}$, $\phi \wedge \psi \in \mathcal{L}$, $\phi \vee \psi \in \mathcal{L}$.*

Our formalism accounts for formulas (beliefs and desires), and the trust degree of information sources. The formalism is a fuzzy extension of that proposed in [6] in two ways: (i) incoming information may be any propositional formula, not only atoms or literals, and (ii) the degree of trust in the sources of information is taken into account.

The first extension makes the formalism more general with respect to previous proposals, in that it allows to express all kinds of incoming information, including disjunctive information.

Thanks to the second extension, 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 the source of the piece of information is. 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].

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-generation rule defined in [6] is defined as follows:

Definition 3 (Desire-Generation Rule) *A desire-generation rule is an expression of the form $R = \{\beta_R, \psi_R \Rightarrow_D^+ d \mid \beta_R, \psi_R \in \mathcal{L}, d \in \{a, \neg a\}, a \in \mathcal{A}\}^2$.*

² The unconditional counterpart of of this rule is $\delta \Rightarrow_D^+ d$ which means that the agent (unconditionally) desires d to degree δ .

Intuitively this means that: "an agent desires d as much as it believes β_R and desires ψ_R . Unlike in most conventional approaches like e.g. [3], in which the authors do not consider at all disjunctive information, here, for sake of simplicity, we make this restriction only for generated desires, i.e. those in the right-hand side of the rules: a generated desire is then represented by a literal.

Given a desire-generation rule R , we shall denote $\text{rhs}(R)$ the literal on the right-hand side of R .

The preferences and habits of the gourmet in the example may be described by means of the following rules:

$$\begin{array}{l} R_1 : \quad \text{ff}, \top \Rightarrow_D^+ \text{hf}, \quad \left| \quad R_4 : \quad \top, \text{hf} \Rightarrow_D^+ \text{ww}, \right. \\ R_2 : \quad \text{ge}, \neg\text{hf} \Rightarrow_D^+ \text{he}, \quad \left| \quad R_5 : \quad \top, \text{hf} \Rightarrow_D^+ \neg\text{hm}, \right. \\ R_3 : \quad \top, \text{hm} \Rightarrow_D^+ \text{rw}, \quad \left| \quad R_6 : \quad 0.7 \Rightarrow_D^+ \text{hm}. \right. \end{array}$$

2.2.1 Agent's State

In this section, we define the mental state of an agent and the semantics of belief and desire formulas.

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 interpretation on \mathcal{A} ;
- \mathcal{R}_J is a set of desire-generation rules, such that, for each desire d , \mathcal{R}_J contains at most one rule of the form $\delta \Rightarrow_D^+ d$;
- \mathcal{J} is a fuzzy set of literals.

\mathcal{B} is the fuzzy interpretation which defines the degree to which the agent believes each atom in \mathcal{A} . Representing the beliefs as a fuzzy interpretation on \mathcal{A} guarantees by construction that the agents beliefs are consistent, i.e., for all atom a we have $\mathcal{B}(a) = 1 - \mathcal{B}(\neg a)$. \mathcal{R}_J contains the rules which generate desires from beliefs and other desires (subdesires). \mathcal{J} contains all literals (positive and negative form of atoms in \mathcal{A}) representing desires which may be deduced from the agents's desire-generation rules. We suppose that an agent can have inconsistent desires, i.e., for each desire d we can have $\mathcal{J}(d) + \mathcal{J}(\neg d) > 1$.

In the gourmet example, your initial state when you step into the restaurant might be described, by $\mathcal{B}(\text{ff}) = 0.2$, $\mathcal{B}(\text{ge}) = 0$, and $\mathcal{J}(\text{rw}) = \mathcal{J}(\text{hm}) = 0.7$, $\mathcal{J}(\text{ww}) = \mathcal{J}(\neg\text{hm}) = \mathcal{J}(\text{hf}) = 0.2$, $\mathcal{J}(\text{he}) = 0$.

By extension, we can compute the truth degree of any belief and desire formulas in \mathcal{L} .

Definition 4 (Degree of fuzzy belief and desires formulas)

Let $\mathcal{S} = \langle \mathcal{B}, \mathcal{J}, \mathcal{R}_J \rangle$ be the state of the agent, $\phi, \psi \in \mathcal{L}$ be formulas. We can extend \mathcal{B} to arbitrary formulas in \mathcal{L} by defining:

$$\mathcal{B}(\top) = 1, \tag{1}$$

$$\mathcal{B}(\perp) = 0, \tag{2}$$

$$\mathcal{B}(\neg\phi) = 1 - \mathcal{B}(\phi), \tag{3}$$

$$\mathcal{B}(\phi \wedge \psi) = \min\{\mathcal{B}(\phi), \mathcal{B}(\psi)\}, \tag{4}$$

$$\mathcal{B}(\phi \vee \psi) = \max\{\mathcal{B}(\phi), \mathcal{B}(\psi)\}. \tag{5}$$

The extension of \mathcal{J} is obtained in the same way, except that Equations 2 and 3 do not hold for \mathcal{J} because \mathcal{J} may be inconsistent. They are replaced by $\mathcal{J}(\perp) = \delta \in [0, 1]$. Besides, if ϕ is a literal, $\mathcal{J}(\phi)$ is directly given by the state of the agent.

Note that since \mathcal{J} need not to be consistent, the De Morgan laws do not hold, in general, for desire formulas.

Definition 5 (Degree of Activation of a Rule) Let R be a desire-generation rule. The degree of activation of R , $\text{Deg}(R)$, is given by

$$\text{Deg}(R) = \min(\mathcal{B}(\beta_R), \mathcal{J}(\psi_R))$$

and for its unconditional counterpart $R = \delta \Rightarrow_D^+ d$: $\text{Deg}(R) = \delta$.

Definition 6 (Degree of Justification) The degree of justification of desire d is defined as $\mathcal{J}(d) = \max_{R \in \mathcal{R}_{\mathcal{J}}: \text{rhs}(R)=d} \text{Deg}(R)$.

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

3 Changes in the Agent's State

The acquisition of a new *consistent* piece of information with a given degree of trust in state \mathcal{S} may cause changes in both degrees of beliefs and justification in the agent's belief and desire sets respectively.

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 piece of information, 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.

We consider the disjunctive normal form (DNF) of the new piece of information β ³, i.e., $\beta = \mathcal{K}_1 \vee \mathcal{K}_2 \vee \dots \vee \mathcal{K}_n$, with $\forall i, \mathcal{K}_i = l_1^i \wedge l_2^i \wedge \dots \wedge l_m^i$, and $l_j^i \in \{a_j^i, \neg a_j^i\}$, with $a_j^i \in \mathcal{A}$.

We suppose that β is consistent. This allows us to dispense with dealing with cases in which inconsistent beliefs make it possible to deduce all formulas and, therefore, to believe everything. If a new piece of information β arrives with degree of trust α , n alternatives are possible: either the agent trusts \mathcal{K}_1 with degree α , or \mathcal{K}_2 with degree α and so on. The value α corresponds to how strongly the agent trusts β . Here, we make a choice, motivated by the Minimal Change Principle [12]. We suppose that the agent chooses the alternative which produces the smallest change in its beliefs. We measure such changes for each disjunct of the incoming formula, thanks to the belief change operator defined below.

Definition 7 (Belief Change Alternatives) Let $a \in \mathcal{A}$ and \mathcal{K}_i be one of the disjuncts of β , the incoming information, whose degree of trust is α . Let \mathcal{B} be the agent's fuzzy belief set. The i th alternative new fuzzy set of beliefs $\mathcal{B}'_i = \mathcal{B} * \frac{\alpha}{\mathcal{K}_i}$ is such that, for all $a \in \mathcal{A}$,

$$\mathcal{B}'_i(a) = \begin{cases} \mathcal{B}(a) \cdot (1 - \alpha) + \alpha, & \text{if } \mathcal{K}_i \models a; \\ \mathcal{B}(a) \cdot (1 - \alpha), & \text{if } \mathcal{K}_i \models \neg a; \\ \mathcal{B}(a), & \text{otherwise.} \end{cases} \quad (6)$$

This operator allows us to update the new degree of the agent's beliefs in each atom $a \in \mathcal{A}$, with respect to both the \mathcal{K}_i disjunct of the incoming information and the trust degree of its source.

Observation 1 If the agent trusts completely ($\alpha = 1$) a source which provides a piece of information confirming (contradicting) a , then the agent will (will not at all) believe a completely (anymore) no matter which its previous degree of belief in a was.

³ This is not a restriction because any well-formed formula of propositional logic has an equivalent DNF expression.

This observation underlines the fact that, in case of a completely trusted source, our operator obeys the Primacy of New Information Principle [8].

We measure the amount of change in beliefs by means of a fuzzy version d_H of the Hamming distance between interpretations: given two fuzzy interpretations \mathcal{I}_1 and \mathcal{I}_2 ,

$$d_H(\mathcal{I}_1, \mathcal{I}_2) = \sum_{a \in \mathcal{A}} |\mathcal{I}_1(a) - \mathcal{I}_2(a)|. \quad (7)$$

As explained previously, based on the minimal change principle, we suppose that the agent chooses the disjunct (or one of the disjuncts in case of tie) with the smallest total amount of change. More formally,

Definition 8 (Belief Change Operator) Let $\beta = \mathcal{K}_1 \vee \dots \vee \mathcal{K}_n$ be the incoming information with trust degree α . The new set of beliefs is given by $\mathcal{B} * \frac{\alpha}{\beta} = \mathcal{B}'_{i^*}$, with

$$i^* = \arg \min_i d_H(\mathcal{B}'_i, \mathcal{B}),$$

where \mathcal{B}'_i is the i th alternative revision as per Definition 7. If there is more than one i such that $d_H(\mathcal{B}'_i, \mathcal{B})$ is minimal, one is chosen arbitrarily.

In the gourmet example, when your friend, whom you trust to a degree $\alpha = 0.8$, tells you “ $ff \vee ge$ ”, you would change your beliefs \mathcal{B} as follows: $\mathcal{K}_1 = ff, \mathcal{K}_2 = ge; \mathcal{B}'_1(ff) = B(ff) \cdot 0.2 + 0.8 = 0.84; \mathcal{B}'_1(ge) = B(ge) = 0; \mathcal{B}'_2(ff) = B(ff) = 0.2; \mathcal{B}'_2(ge) = B(ge) \cdot 0.2 + 0.8 = 0.8; d_H(\mathcal{B}'_1, \mathcal{B}) = 0.64; d_H(\mathcal{B}'_2, \mathcal{B}) = 0.8; \text{ therefore, } i^* = 1 \text{ and } \mathcal{B} * (ff \vee ge) = \mathcal{B}'_1$.

Proposition 1 If $\mathcal{K}_{i^*} \models a$, i.e., \mathcal{K}_{i^*} confirms a to a certain degree, then applying the operator $*$ never causes the belief degree of a to decrease, i.e., $\mathcal{B}'(a) \geq \mathcal{B}(a)$.

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

Proposition 2 If $\mathcal{K}_{i^*} \models \neg a$, i.e., \mathcal{K}_{i^*} contradicts a to a certain degree, then applying the operator $*$ never causes the belief degree of a to increase, i.e., $\mathcal{B}'(a) \leq \mathcal{B}(a)$.

Proof: We have $\mathcal{B}'(a) - \mathcal{B}(a) = -\alpha \cdot \mathcal{B}(a) \leq 0$. \square

The semantics of our belief change operator is defined by the following properties. Here \mathcal{B} represents a fuzzy belief set, β the incoming trusted information with degree of trust α , supp is the support of a fuzzy set, and \cup, \cap, \subseteq and \supseteq are fuzzy operators.

- **(P * 1)(Stability)** The result of applying $*$ in \mathcal{B} with β is always a fuzzy set of beliefs: $\mathcal{B} * \frac{\alpha}{\beta}$ is a fuzzy set of beliefs.
- **(P * 2)(Expansion)** If \mathcal{K}_{i^*} contains only positive atoms, then the fuzzy set \mathcal{B} expands: $\text{supp}(\mathcal{B} * \frac{\alpha}{\beta}) \supseteq \text{supp}(\mathcal{B})$.
- **(P * 3)(Shrinkage)** If \mathcal{K}_{i^*} contains only negated atoms, then the fuzzy set \mathcal{B} shrinks: $\text{supp}(\mathcal{B} * \frac{\alpha}{\beta}) \subseteq \text{supp}(\mathcal{B})$.
- **(P * 4)(Invariance)** If the new information is completely untrusted, i.e., $\alpha = 0$, invariance holds: $(\alpha = 0) \Rightarrow (\mathcal{B} * \frac{\alpha}{\beta} = \mathcal{B})$.
- **(P * 5)(Predictability)** The result of applying $*$ contains all belief atoms in $\text{supp}(\mathcal{B} \cup \{\frac{\alpha}{\beta}\})$: $\text{supp}(\mathcal{B} * \frac{\alpha}{\beta}) \supseteq \text{supp}(\mathcal{B} \cup \{\frac{\alpha}{\beta}\})$.
- **(P * 6)(Identity)** The result of applying $*$ does not depend on the particular information. If $\beta_1 \equiv \beta_2$ and $\alpha_1 = \alpha_2$: $\mathcal{B} * \frac{\alpha_1}{\beta_1} = \mathcal{B} * \frac{\alpha_2}{\beta_2}$.

When all beliefs are crisp and the trust in new information is complete ($\alpha = 1$), our fuzzy belief-change operator satisfies the six basic AGM revision rationality postulates K*1–K*6 [10]. In order to show that, let us consider the standard definition of expansion of a crisp set of formulas \mathcal{B} with a formula $\phi \in \mathcal{L}$ as $\mathcal{B} + \phi = \{\psi : \mathcal{B} \cup \{\phi\} \vdash \psi\}$.

Proposition 3 *If \mathcal{B} is crisp, ϕ is new information whose trust is $\alpha = 1$, the following hold:*

1. $\mathcal{B}' = \mathcal{B} * \phi$ is a crisp interpretation (K*1);
2. $\mathcal{B}'(\phi) = 1$ (K*2);
3. $\mathcal{B}' \subseteq \mathcal{B} + \phi$ (K*3);
4. if $\mathcal{B}(\neg\phi) = 0$, then $\mathcal{B} + \phi \subseteq \mathcal{B}'$ (K*4);
5. if $\phi \equiv \psi$, then $\mathcal{B} * \phi = \mathcal{B} * \psi$; (K*6);

For the convenience of the reader, the corresponding AGM rationality postulate has been indicated between parentheses for each thesis. Note that Postulate K*5, which in our formalism would be “ $\mathcal{B} * \phi = \mathcal{L}$ iff $\phi = \perp$ ”, is not relevant to our discussion, since we have made the assumption that new information is never inconsistent; therefore, it has not been considered.

Proof: To prove Thesis 1, we observe that, when $\alpha = 1$, for all atom a , $\mathcal{B}'(a) \in \{0, \mathcal{B}(a), 1\}$; but $\mathcal{B}(a) \in \{0, 1\}$, since \mathcal{B} is crisp; therefore, $\mathcal{B}'(a) \in \{0, 1\}$ as well.

To prove Thesis 2, let us consider \mathcal{K}_{i^*} , the chosen alternative disjunct of ϕ : a sufficient condition for $\mathcal{B}'(\phi) = 1$ is that $\mathcal{B}'(\mathcal{K}_{i^*}) = 1$; now, $\mathcal{K}_{i^*} = l_1^* \wedge \dots \wedge l_m^*$; it is easy to verify that, according to Definition 7, $\mathcal{B}'(l_i^*) = 1$ for all $i = 1, \dots, m$; therefore, $\mathcal{B}'(\mathcal{K}_{i^*}) = \min_i \{\mathcal{B}'(l_i^*)\} = 1$ and the thesis follows.

As for Thesis 3, it follows trivially if $\mathcal{B} + \phi = \mathcal{L}$, i.e., if $\mathcal{B}(\neg\phi) = 1$. In all other cases, i.e., when $\mathcal{B}(\neg\phi) = 0$, we have $\mathcal{B}(\phi) = 1$, and, because of the minimal change principle, $\mathcal{B}' = \mathcal{B}$, which proves the thesis.

The proof of Thesis 4 is similar: $\mathcal{B}(\neg\phi) = 0$ implies $\mathcal{B}(\phi) = 1$, whence one concludes $\mathcal{B}' = \mathcal{B}$; furthermore, $\mathcal{B} + \phi = \mathcal{B}$, which verifies the thesis.

Finally, to prove Thesis 5, we recall that, if $\phi \equiv \psi$, their DNFs are identical; therefore, $\mathcal{B} * \phi = \mathcal{B} * \psi$ by definition. \square

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 β , be a new belief trusted to degree α , ($\frac{\alpha}{\beta}$). To account for the changes in the desire set caused by this new acquisition, we have to recursively:

- (i) calculate for each rule $R \in \mathcal{R}_{\mathcal{J}}$ their new activation degree by considering \mathcal{B}' and
- (ii) update the justification degree of all desires in their right-hand side ($\text{rhs}(R)$).

The new desire set \mathcal{J}' is obtained by executing the algorithm in Figure 1 with the following inputs: $\mathcal{B}' = \mathcal{B} * \frac{\alpha}{\beta}$, $\mathcal{R}_{\mathcal{J}}$, and \mathcal{A} . 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., $\forall d \in \{a, \neg a\}$, with $a \in \mathcal{A}$, $d \in C_k \Rightarrow \mathcal{J}'_k(d) \neq \mathcal{J}'_{k-1}(d)$. Step 1 updates \mathcal{B}' with respect to the incoming information $\frac{\alpha}{\beta}$, and initializes to empty, the set of desires such that justification degrees directly changes with the arrival of $\frac{\alpha}{\beta}$, C_0 . Step 2 updates C_0 . Steps 3 and 4

1. $\mathcal{B}' \leftarrow \mathcal{B} * \frac{\alpha}{\beta}$; $k \leftarrow 1$; $C_0 \leftarrow \emptyset$;
2. For each $d \in \{a, \neg a\}$ with $a \in \mathcal{A}$ 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}' ;
 - (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 $\psi_{R_j} \models d$ do
 - A. calculate $\text{Deg}(R_j)$ 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. $k \leftarrow k + 1$.
- until $C_{k-1} = \emptyset$.
4. for all d , $\mathcal{J}'(d)$ is given by the following equation:

$$\mathcal{J}'(d) = \begin{cases} \mathcal{J}(d), & \text{if } d \notin C; \\ \mathcal{J}'_k(d), & \text{otherwise,} \end{cases} \quad (8)$$

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 1. An algorithm to compute the new desire set upon arrival of a new belief.

update desire degrees which are indirectly changed by the incoming information.

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

In the gourmet example, learning $\beta = ff \vee ge$ with $\alpha = 0.8$, which has you change your beliefs to \mathcal{B}' such that $\mathcal{B}'(ff) = 0.84$ and $\mathcal{B}'(ge) = 0$, makes \mathcal{J} change to \mathcal{J}' such that $\mathcal{J}'(ww) = \mathcal{J}'(hf) = 0.84$, $\mathcal{J}'(he) = 0$, $\mathcal{J}'(rw) = \mathcal{J}'(hm) = 0.7$, and $\mathcal{J}'(\neg hm) = 0.84$.

Proposition 4 *If the chosen disjunct \mathcal{K}_{i^*} does not contain negated atoms, then $\mathcal{J}' = \bigcup_{k=0}^{\infty} \mathcal{J}'_k$.*

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

Proposition 5 *If the chosen disjunct \mathcal{K}_{i^*} only contains negated atoms, then $\mathcal{J}' = \bigcap_{k=0}^{\infty} \mathcal{J}'_k$.*

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

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-generation 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 $\delta \Rightarrow_D^+ d$ into $\mathcal{R}_{\mathcal{J}}$, resulting in the new base $\mathcal{R}'_{\mathcal{J}}$. By definition of a desire-generation rule base, $\mathcal{R}'_{\mathcal{J}}$ must not contain a $\delta' \Rightarrow_D^+ d$ with $\delta \neq \delta'$. How does \mathcal{S} change with the arising of the new desire $\frac{\delta}{d}$?

1. if $\{\delta' \Rightarrow_D^+ d\} \in \mathcal{R}_{\mathcal{J}}$ then $\mathcal{R}'_{\mathcal{J}} \leftarrow (\mathcal{R}_{\mathcal{J}} \setminus \{\delta' \Rightarrow_D^+ d\}) \cup \{\delta \Rightarrow_D^+ d\}$; else $\mathcal{R}'_{\mathcal{J}} \leftarrow \mathcal{R}_{\mathcal{J}} \cup \{\delta \Rightarrow_D^+ d\}$;
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 $\psi_{R_j} \models d$ 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. for all d , $\mathcal{J}'(d)$ is given by Equation 8.

Figure 2. An algorithm to compute the new desire set upon the arisal of a new desire.

- Any rule $\delta' \Rightarrow_D^+ d$ with $\delta \neq \delta'$ is retracted from $\mathcal{R}_{\mathcal{J}}$,
- $\delta' \Rightarrow_D^+ d$ is added to $\mathcal{R}_{\mathcal{J}}$,

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 2.

4 Goal Adoption

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 problem \mathcal{P} to solve, not all generated desires can be adopted at the same time, especially when they are not feasible at the same time. We assume we dispose of a \mathcal{P} -dependent function $\mathcal{F}_{\mathcal{P}}$ which, 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 feasible solution found. We may call γ the *degree of feasibility* of \mathcal{J} given \mathcal{B} , i.e., $\mathcal{F}_{\mathcal{P}}(\mathcal{B}, \mathcal{J}) = \gamma$.

Definition 9 (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} is justified: $\mathcal{G} \subseteq \mathcal{J}$, i.e., $\forall d \in \{a, \neg a\}, a \in \mathcal{A}, \mathcal{G}(d) \leq \mathcal{J}(d)$;
2. \mathcal{G} is γ -feasible: $\mathcal{F}_{\mathcal{P}}(\mathcal{B}, \mathcal{G}) \geq \gamma$;
3. \mathcal{G} is consistent: $\forall d \in \{a, \neg a\}, a \in \mathcal{A}, \mathcal{G}(d) + \mathcal{G}(\neg d) \leq 1$.

In the gourmet example, \mathcal{J}' is inconsistent, in that $\mathcal{J}'(hm) + \mathcal{J}'(\neg hm) = 1.54 > 1$; on the other hand, consistency requires that $\mathcal{G}(hm) + \mathcal{G}(\neg hm) \leq 1$; therefore, one possible choice for \mathcal{G} could be such that $\mathcal{G}(hm) = 0.45$ and $\mathcal{G}(\neg hm) = 0.55$, or even $\mathcal{G}(hm) = 0$ and $\mathcal{G}(\neg hm) = 0.84$.

In general, given a fuzzy set of desires \mathcal{J} , there may be more than one possible γ -goal set \mathcal{G} . However, a rational agent in state $\mathcal{S} = \langle \mathcal{B}, \mathcal{J}, \mathcal{R}_{\mathcal{J}} \rangle$, for practical reasons, may need to elect one precise set of goals, \mathcal{G}^* , to pursue, which depends on \mathcal{S} . The choice of one γ -goal set over the others may be based on a preference relation \succeq on

desire sets, as proposed in [7], where it is required that a goal election function G_{γ} is such that:

- $\forall \mathcal{S}, G_{\gamma}(\mathcal{S})$ is a γ -goal set, i.e., it does indeed return a γ -goal set; and
- $\forall \mathcal{S}$, if \mathcal{G} is a γ -goal set, then $G_{\gamma}(\mathcal{S}) \succeq \mathcal{G}$, i.e., the γ -goal set returned by function G_{γ} and then adopted by the agent is “optimal”.

The issue of defining a specific goal election function is a critical part of constructing a rational agent framework. Such issue falls out of the scope of this work.

5 Summary

We have investigated how trust in a source of information can influence the degree of an agent’s beliefs, and how these graded beliefs influence the agent’s generated desires and then its adopted goals. We propose a new fuzzy belief change operator to deal with this new kind of information and two algorithms for updating the agent’s desire set after the arrival of a new, even partially trusted, piece of information, and a new unconditional desire. Finally, requirements for goal adoption have been stated.

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*, pp. 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*, pp. 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 and A. Tettamanzi, ‘Towards a framework for goal revision’, in *Proceedings of BNAIC’06*, pp. 99–106, (2006).
- [7] C. da Costa Pereira and A. Tettamanzi, ‘Goal generation with relevant and trusted beliefs’, in *Proceedings of AAMAS’08*, pp. 397–404, (2008).
- [8] M. Dalal, ‘Investigations into a theory of knowledge base revision’, in *AAAI*, pp. 475–479, (1988).
- [9] 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).
- [10] P. Gärdenfors, ‘Belief revision: A vademecum’, in *Meta-Programming in Logic*, 1–10, Springer, Berlin, (1992).
- [11] S. O. Hansson, ‘Ten philosophical problems in belief revision’, *Journal of Logic and Computation*, **13**(1), 37–49, (February 2003).
- [12] H. Katsuno and A. O. Mendelzon, ‘Propositional knowledge base revision and minimal change’, *Artif. Intell.*, **52**(3), 263–294, (1991).
- [13] S. Parsons and P. Giorgini, ‘An approach to using degrees of belief in BDI agents’, in *Information, Uncertainty, Fusion*, eds., B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, Kluwer, Dordrecht, (1999).
- [14] A. S. Rao and M. P. Georgeff, ‘Modeling rational agents within a BDI-architecture’, in *Proceedings of KR’91*, pp. 473–484, (1991).
- [15] S. Shapiro, Y. Lespérance, and H. J. Levesque, ‘Goal change’, in *Proceedings of IJCAI’05*, pp. 582–588, (2005).
- [16] J. Thangarajah, L. Padgham, and J. Harland, ‘Representation and reasoning for goals in BDI agents’, in *Proceedings of CRPITS’02*, pp. 259–265, (2002).
- [17] R. H. Thomason, ‘Desires and defaults: A framework for planning with inferred goals’, in *Proceedings of KR’00*, pp. 702–713, (2000).
- [18] M. Birna van Riemsdijk, *Cognitive Agent Programming: A Semantic Approach*, Ph.D. dissertation, University of Utrecht, 2006.
- [19] L. A. Zadeh, ‘Fuzzy sets’, *Information and Control*, **8**, 338–353, (1965).