An Integrated Possibilistic Framework for Goal Generation in Cognitive Agents

Célia da Costa Pereira
Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano
Via Bramante 65, I-26013 Crema (CR), Italy
celia.pereira@unimi.it

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

ABSTRACT

We propose an integrated theoretical framework, grounded in possibility theory, to account for all the aspects involved in representing and changing beliefs, representing and generating justified desires, and selecting goals based on current and uncertain beliefs about the world, and the preferences of the agent.

Beliefs and desires of a cognitive agent are represented as (two distinct) possibility distributions. This is the original part of the proposed framework in the sense that there does not really exist a full-fledged possibilistic approach to BDI. In the proposed framework: (i) the possibility distribution representing the qualitative utilities associated with desires are the result of a rule-based deliberative process, that is, they depend on the mental state of the agent and are not given *a priori*; and (ii) the criteria for choosing goals take into account not only the fact that goals must be consistent in the logical sense, but also in the cognitive sense.

Categories and Subject Descriptors

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

General Terms

Theory

Keywords

Beliefs, desires and goals, fuzzy logic, possibility theory

1. INTRODUCTION AND RELATED WORK

There is a consensus among researchers that the generation of the goals to be adopted by an agent depends on its mental state [2, 7, 10, 12]. It can be the result of an explicit external request that is *accepted* by the agent, e.g., [26]; or a *consequence* of the agent's mental attitudes, e.g., [10]. Instead, the choice of the *best* set of goals to be adopted (pursued) depends also on the *consistency* (or *feasibility*) of such goals. A common assumption in the agent theory literature states that "achievement goals that are believed to be impossible to achieve should be dropped" [11, 25].

There are two directions followed by the researchers to define desire/goal consistency. The one that considers the steps of both goal generation and adoption as a whole, and the one that considers them separately. In the former case, the evaluation of the consistency of a desire set takes the cognitive components into account, but can lead to a goal set that is not the (or one among the) maximal consistent set. In the latter case, the evaluation of consistency does not take the cognitive components of the agent into account (logical consistency). This can lead the agent to choose sets of desires which are logically consistent but inconsistent from the cognitive point of view.

We propose a possibilistic approach in which the generation and the adoption parts are considered separately and propose a new and possibilistic-based definition of desire/goal consistency which incorporates the two points of view. Using a possibilistic framework to represent beliefs and desires allows one to also represent *partially sure beliefs* and *partiallly desirable* world states. The originality of what we are proposing with respect to the existing works is the use possibility distributions to represent beliefs and desires in BDI agents:

(i) the possibility distribution representing the qualitative utilities associated to desires are the result of a deliberative process, that is, it depends on the mental state of the agent;

(ii) we consider that desires may be inconsistent and propose a way to calculate the degree of (logical and cognitive) consistency of the agent's desires;

(iii) we make an explicit distinction between *goals* and *desires*;

(iv) with respect to, e.g., [7, 13], we extend the representation of desires to arbitrary formulas and not just literals.

The first step was the choice of the most suitable representation for beliefs. Recently, we have proposed an approach considering that belief is a matter of degree and, as a consequence, also goals are adopted (desired) to some extent [12, 13]. In such approach, we basically dealt with a classical (Boolean) propositional language, but allowed for the definition of graded beliefs and, for that purpose, we defined a sort of truth-functional fuzzy semantics for beliefs (and desires). Here, we propose a possibilistic approach which is more suitable to representing uncertain beliefs. Indeed, there is a main difference between truth degrees and degrees of uncertainty, made clear by the following example, due to Bezdek and Pal [19]. In terms of binary truth-values, a bottle is viewed as full or empty. If one accounts for the quantity of liquid in the bottle, one may say the bottle is "half full" for instance. In that case, "full" can be seen as a fuzzy predicate and the degrees of truth of "The bottle is full" reflects the amount of liquid in the bottle. The situation is quite different when expressing our ignorance about whether the bottle is full or empty (given that we know only one of the two situations is the true one). This reflects our degree of evidence to the fact that

Cite as: An Integrated Possibilistic Framework for Goal Generation in Cognitive Agents, Célia da Costa Pereira and Andrea G. B. Tettamanzi, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1239-1246

the bottle is full or empty. To say that the possibility (or the probability) that the bottle is full is 1/2 does not mean that the bottle is half full.

We propose to adopt the representation of beliefs as uncertain, in the sense that, given a piece of information, and given the agent's current knowledge, the agent can assert how sure it is about that piece of information. The agent's beliefs are then represented thanks to a possibility distribution and we adapt the belief conditioning operator proposed by Dubois and Prade [18], and propose an extension of the AGM-like revision postulates proposed in [18] to the case of partially sure new information.

The second step was the choice of a suitable representation for desires. A consequence of representing beliefs as a matter of degree is that desires also have to be considered as such.

In [22, 23] the authors start from Boutilier's preference-based handling of conditional desires [6, 27] (if a then ideally b) for modeling preferences and suggest to view such conditional desires as constraints on utility functions. The violation of such contraints induces a loss of utility, while their satisfaction induces a gain of utility. Dastani et al. [15], who compare BDI systems to Qualitative Decision Theory, made an analysis about the profit derived from the synergy of both systems. They proposed a decision-theoretic approach where losses and gains are associated with rules. While our approach is purely qualitative and possibilistic, in the sense that only the ordering between positive desires is considered, the above mentioned approaches are based on additive utilities.

There are several other approaches which deal with and represent graded desires [3, 5, 9, 24]. In particular, there are recent interesting works which have pointed out the importance of making a distinction between positive and negative preferences. Negative preferences correspond to what is rejected, considered unacceptable, while positive preferences correspond to what is desired. But what is tolerated (i.e., not rejected) is not necessarily desired. Such bipolar preferences can be represented in possibilistic logic [14] by two separate sets of formulas: prioritized constraints, which describe what is more or less tolerated, and weighted positive preferences, expressing what is particularly desirable. In none of these approaches, the degrees of satisfaction and tolerance of preferences are supposed to derive from the cognitive aspects of the agent. Most approaches assume that they are given *a priori*.

The paper is organized as follows: Section 2 provides an overall description of the proposed framework, whose details are given in the rest of the paper; Section 3 provides essential background on fuzzy sets and possibility theory, and introduces a possibilistic representation of beliefs, desires, and the mental state of an agent. After these preliminaries, Section 4 discusses how beliefs entertained by an agent may change in response to the receipt of new information from the environment; Section 5 models the deliberation mechanisms whereby an agent dynamically generates its desires based on its mental state, while Section 6 describes how goals are selected from the agent desires in a rational way. Section 7 concludes.

2. OVERVIEW OF THE FRAMEWORK

A schematic illustration of the proposed framework is provided in Figure 1 as a handy navigation tool.

The framework may be classified as a BDI model of agency. The three distinct layers of beliefs, desires, and intentions are boxed in gray, as is the perimeter of the agent.

The agent interacts with the world by receiving information ϕ from one or more *sources*, and by performing *actions*. Furthermore, the agent is "programmed" by its user (or owner) by inserting one or more desire-generation rules r into rule-base $\mathcal{R}_J$.

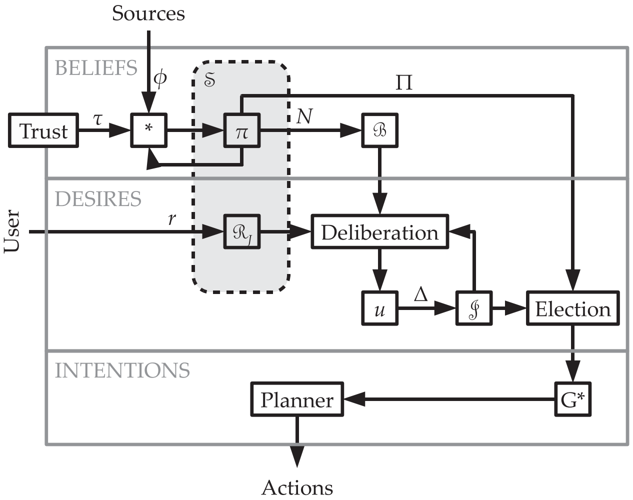

Figure 1: A schematic illustration of the proposed BDI framework. The meaning of the symbols is explained in the text.

The agent has an internal mental state $\mathcal{S}$ that is completely described by a possibility distribution π, representing beliefs, and by a set of desire-generation rules $\mathcal{R}_J$. Possibility distribution π is dynamic and changes as new information ϕ is received from a source. A *trust* module, whose details are not covered in this paper, assigns a trust degree τ to each source. A belief change operator $*$ changes π in light of new information ϕ while taking the degree τ to which the source of ϕ is trusted into account. Possibility distribution π induces an explicit representation $\mathcal{B}$ of the agent's beliefs as a necessity measure N.

The set $\mathcal{J}$ of the agent's justified desires is generated dynamically through a deliberation process which applies the rules in $\mathcal{R}$ to the current beliefs and desires to produce a possibility distribution of qualitative utility u, which induces $\mathcal{J}$ as a guaranteed possibility Δ.

Finally, the agent rationally *elects* its goals G^* from the justified desires $\mathcal{J}$ as the most desirable of the possible sets of justified desires, according to a possibility measure Π induced by π. The agent then plans its actions to achieve the elected goals G^* by means of a planner module, whose discussion lies outside of the scope of this paper.

3. POSSIBILISTIC REPRESENTATION

In this section, we introduce a possibilistic representation of beliefs and desires, and define the mental state of an agent. Essential background and definitions on fuzzy set theory and possibility theory are given.

3.1 Fuzzy Sets

Fuzzy sets [28] are a generalization of classical (crisp) 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 2. Given a fuzzy set

A, its *core* is the (conventional) set of all elements x such that $A(x) = 1$; its *support*, $\mathrm{supp}(A)$, 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.

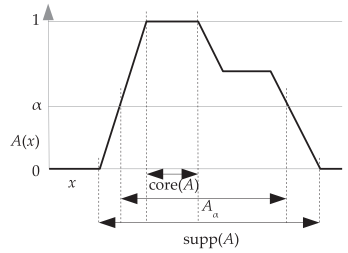

Figure 2: Core, support, and α-cuts of a set A of the real line.

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,

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

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

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

Finally, given two fuzzy sets A and B, $A \subseteq B$ if and only if, for all element x, $A(x) \leq B(x)$.

3.2 Possibility Theory

The membership function of a fuzzy set describes the more or less possible and mutually exclusive values of one (or more) variable(s). Such a function can then be seen as a possibility distribution [29]. Indeed, if F designates the fuzzy set of possible values of a variable X, $\pi_X = \mu_F$ is called the possibility distribution associated to X. The identity $\mu_F(v) = \pi_X(v)$ means that the membership degree of v to F is equal to the possibility degree of X being equal to v when all we know about X is that its value is in F. A possibility distribution for which there exists a completely possible value ($\exists v_0; \pi(v_0) = 1$) is said to be *normalized*.

Definition 1 (Possibility and Necessity Measures). *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:*

$$\Pi(A) = \max_{s \in A} \pi(s); \quad (4)$$

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

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 $\bar{A}$.

Another interesting measure that can be defined based on a possibility distribution is *guaranteed possibility* [20].

Definition 2 (Guaranteed Possibility Measure). *Given a possibility distribution π, a guaranteed possibility measure, noted Δ, is defined as:*

$$\Delta(A) = \min_{s \in A} \pi(s); \quad (6)$$

In words, the guaranteed possibility measure estimates to what extent *all* the values in A are actually possible according to what is known, i.e., any value in A is at least possible at degree $\Delta(A)$.

A few properties of possibility, necessity, and guaranteed possibility measures induced by a normalized possibility distribution on a finite universe of discourse Ω are the following. For all subsets $A, B \subseteq \Omega$:

1. $\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\}$;
2. $\Pi(\emptyset) = N(\emptyset) = 0, \quad \Pi(\Omega) = N(\Omega) = 1$;
3. $N(A \cap B) = \min\{N(A), N(B)\}$;
4. $\Pi(A) = 1 - N(\bar{A})$ (duality);
5. $N(A) \leq \Pi(A)$;
6. $N(A) > 0$ implies $\Pi(A) = 1$;
7. $\Pi(A) < 1$ implies $N(A) = 0$;
8. $\Delta(A) \leq \Pi(A)$.

A consequence of these propoerties is that $\max\{\Pi(A), \Pi(\bar{A})\} = 1$. In case of complete ignorance on A, $\Pi(A) = \Pi(\bar{A}) = 1$.

3.3 Negative and Positive Information

A relatively recent work on cognitive psichology [8] pointed out that *negative information* and *positive information* are processed separately in the brain. This strengthens the idea that negative and positive information require different representation models and different reasoning techniques. Here, like for example in [20], by negative information we mean information that *restricts* the number of situations deemed possible, by just keeping what is possible because not ruled out by the available knowledge. This is in line with the classical view in logic according to which each new piece of information declares some worlds impossible. Positive information instead, leads to a disjunctive accumulation of information in the sense that the more one is informed (with positive information or facts), the larger the range of worlds which are guaranteed to be possible.

When modeling knowledge (or beliefs), such a bipolar view is suited to making a distinction between what is possible because it is consistent with the available information, and what is possible for sure because observed by facts. Let us consider the following example about bipolar knowledge proposed by Dubois and Prade in [20]. Assume for instance one has some information about the opening hours and the entrance fee of a museum M. We may know that museum M is open from 2pm to 4pm, and certainly closed at night (from 9pm to 9am). Note that nothing forbids museum M to be open in the morning although there is no positive evidence supporting it. Its entrance fee is neither less than 2 euros nor more than 8 euros (following legal regulations), prices between 4 and 5 euros are really possible (they are prices actually proposed by similar museums).

When modeling desires (or preferences), the bipolar view is suited to distinguishing between *positive desires* (positive preferences) which are associated to *satifaction degrees* and *negative preferences* which reflect *what is not rejected as unsatisfactory* and are associated to *tolerance degrees*. For example, a teacher can express his preferences about the days of the week in which he would prefer to teach a class. He can express two kind of preferences. The positive ones with a list of preferred days, each one associated to a satisfaction degree; and the list of negative preferences with the days which are unacceptable for him to a certain degree of tolerance.

Here, we adopt the representation of beliefs proposed in [18] but, besides, we propose a belief updating operator which is an adaptation of Dubois and Prade's belief conditioning operator for belief revision in light of the acquisition of partially sure *negative information*.

For sake of simplicity, we just consider the positive side of desires. Therefore, our agent just generates desires which it would like to be satisfied.

3.4 Language and Interpretations

Information manipulated by a cognitive agent must be represented symbolically. To develop our theoretical framework, we adopt perhaps the simplest symbolic representation, in the form of a classical propositional language.

DEFINITION 3 (LANGUAGE). *Let $\mathcal{A}$ be a* finite[1] *set of atomic propositions and let $\mathcal{L}$ be the propositional language such that $\mathcal{A} \cup \{\top, \bot\} \subseteq \mathcal{L}$, and, $\forall \phi, \psi \in \mathcal{L}$, $\neg\phi \in \mathcal{L}$, $\phi \wedge \psi \in \mathcal{L}$, $\phi \vee \psi \in \mathcal{L}$.*

As usual, one may define additional logical connectives and consider them as useful shorthands for combinations of connectives of $\mathcal{L}$, e.g., $\phi \supset \psi \equiv \neg\phi \vee \psi$.

We will denote by $\Omega = \{0,1\}^{\mathcal{A}}$ the set of all interpretations on $\mathcal{A}$. An interpretation $\mathcal{I} \in \Omega$ is a function $\mathcal{I} : \mathcal{A} \to \{0,1\}$ assigning a truth value $p^{\mathcal{I}}$ to every atomic proposition $p \in \mathcal{A}$ and, by extension, a truth value $\phi^{\mathcal{I}}$ to all formulas $\phi \in \mathcal{L}$.

DEFINITION 4. *The notation $[\phi]$ denotes the set of all models (namely, interpretations satisfying ϕ) of a formula $\phi \in \mathcal{L}$:*

$$[\phi] = \{\mathcal{I} \in \Omega : \mathcal{I} \models \phi\}.$$

Likewise, if $S \subseteq \mathcal{L}$ is a set of formulas,

$$[S] = \{\mathcal{I} \in \Omega : \forall\phi \in S, \mathcal{I} \models \phi\} = \bigcap_{\phi \in S} [\phi].$$

3.5 Representing Beliefs and Desires

A possibility distribution π can represent a complete preorder on the set of possible interpretations $\mathcal{I} \in \Omega$. This is the reason why, intuitively, at a semantical level, a possibility distribution can either represent the available knowledge (or beliefs) of an agent, or its preferences (or desires). When representing knowledge, $\pi(\mathcal{I})$ acts as a restriction on possible interpretations and represents the degree of compatibility of interpretation $\mathcal{I}$ with the available knowledge about the real world. When representing desires, $\pi(\mathcal{I})$ corresponds to the degree to which the agent would be satisfied (i.e., content, pleased, demanding no more, considering it enough, *not* in the logical sense) if the world were the one described by interpretation $\mathcal{I}$. To avoid confusion, in this latter case, we will write $u(\mathcal{I})$ and call u a *qualitative utility*, because this best reflects its intended meaning and purpose; of course, it will be understood that, formally, $u(\cdot)$ is a possibility distribution. By convention, $\pi(\mathcal{I}) = 1$ means that it is totally possible for $\mathcal{I}$ to be the real world. In the case of desires, $u(\mathcal{I}) = 1$ means that $\mathcal{I}$ is fully satisfactory. Instead, $0 < \pi(\mathcal{I}) < 1$ (or $0 < u(\mathcal{I}) < 1$) means that $\mathcal{I}$ is only somewhat possible (or satisfactory), while $\pi(\mathcal{I}) = 0$ (or $u(\mathcal{I}) = 0$) means that $\mathcal{I}$ is certainly not the real world (or not especially satisfactory in case of preferences, though this by no way means rejection). Interpretation $\mathcal{I}$ is more possible (preferred) than interpretation $\mathcal{I}'$ when $\pi(\mathcal{I}) > \pi(\mathcal{I}')$ ($u(\mathcal{I}) > u(\mathcal{I}')$).

We represent the beliefs and desires of a cognitive agent thanks to two different possibility distributions π and u respectively. Thus, a normalized possibility distribution π means that there exists at least one possible situation which is consistent with the available knowledge. All these considerations are in line with what proposed in [3, 20].

In the following, the formal representation of beliefs and desires is proposed.

3.5.1 Representing Desires

Following a common assumption in the agent theory literature, in our framework the desires of an agent depend on its beliefs. Therefore, while we propose that desires be represented by means of a possibility distribution of qualitative utility, one must understand that such a distribution is just an epiphenomenon of an underlying, more primitive mechanism which determines how desires arise. A suitable description of such mechanism may be given in terms of desire-generation rules.

DEFINITION 5 (DESIRE-GENERATION RULE). *A desire-generation rule R is an expression of the form $\beta_R, \psi_R \Rightarrow_D^+ \phi$*[2]*, where $\beta_R, \psi_R, \phi \in \mathcal{L}$. The unconditional counterpart of this rule is $\alpha \Rightarrow_D^+ \phi$, with $\alpha \in (0,1]$.*

The intended meaning of a conditional desire-generation rule is: "an agent desires every world in which ϕ is true at least as much as it believes β_R and desires ψ_R", or, put in terms of qualitative utility, "the qualitative utility attached by the agent to every world satisfying ϕ is greater than, or equal to, the degree to which it believes β_R and desires ψ_R". The intended meaning of an unconditional rule is that the qualitative utility of every world $\mathcal{I} \models \phi$ is at least α for the agent.

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

3.5.2 Representing Graded Beliefs

As convincingly argued by Dubois and Prade [20], a *belief*, which is a component of an agent's cognitive state, can be regarded as a necessity degree induced by a normalized possibility distribution π on the possible worlds $\mathcal{I}$:

$$\pi : \Omega \to [0,1]; \tag{7}$$

where $\pi(\mathcal{I})$ is the possibility degree of interpretation $\mathcal{I}$. It represents the plausibility order of the possible world situation represented by interpretation $\mathcal{I}$.

DEFINITION 6 (GRADED BELIEF). *Let N be the necessity measure induced by π, and ϕ be a formula. The degree to which the agent believes ϕ is given by:*

$$\mathcal{B}(\phi) = N([\phi]) = 1 - \max_{\mathcal{I} \not\models \phi}\{\pi(\mathcal{I})\}. \tag{8}$$

Straightforward consequences of the properties of possibility and necessity measures are that $\mathcal{B}(\phi) > 0 \Rightarrow \mathcal{B}(\neg\phi) = 0$, this means that if the agent somehow believes ϕ then it cannot believe $\neg\phi$ at all; and

$$\begin{aligned}
\mathcal{B}(\top) &= 1, & (9)\\
\mathcal{B}(\bot) &= 0, & (10)\\
\mathcal{B}(\phi \wedge \psi) &= \min\{\mathcal{B}(\phi), \mathcal{B}(\psi)\}, & (11)\\
\mathcal{B}(\phi \vee \psi) &\geq \max\{\mathcal{B}(\phi), \mathcal{B}(\psi)\}. & (12)
\end{aligned}$$

[1] Like in [4], we adopt the restriction to the finite case in order to use standard definitions of possibilistic logic. Extensions of possibilistic logic to the infinite case are discussed for example in [16].

[2] Note that the implication used to define a desire-generation rule is not the material implication.

3.6 Mental State

We now have the elements to define the mental state of an agent, which consists of its beliefs and the rules defining the deliberation mechanism whereby desires are generated based on beliefs.

DEFINITION 7 (MENTAL STATE). *The state of an agent is completely described by a pair $\mathcal{S} = \langle \pi, \mathcal{R}_J \rangle$, where*

- *π is a possibility distribution which induces the agent's beliefs $\mathcal{B}$;*
- *$\mathcal{R}_J$ is a set of desire-generation rules which, together with $\mathcal{B}$, induce a qualitative utility assignment u.*

4. BELIEFS

In this section, we propose a belief change operator which allows to update the possibility distribution π in light of new trusted negative information when the hypothesis of *purely inertial* world made in [18] is weakened, in that an agent may safely assume the world has not change until it receives evidence to the contrary and, in that case, the change is specific to the evidence received.

Here, we suppose that a source of information may be considered trusted to a certain extent. This means that its membership degree to the fuzzy set of trusted sources is $\tau \in [0, 1]$. Let $\phi \in \mathcal{L}$ be incoming information from a source trusted to degree τ. The belief change operator is defined as follows:

DEFINITION 8 (BELIEF CHANGE OPERATOR). *The possibility distribution π' which induces the new belief set $\mathcal{B}'$ after receiving information ϕ is computed from possibility distribution π relevant to the previous belief set $\mathcal{B}$ ($\mathcal{B}' = \mathcal{B} * \frac{\tau}{\phi}$, $\pi' = \pi * \frac{\tau}{\phi}$) as follows: for all interpretation $\mathcal{I}$,*

$$\pi'(\mathcal{I}) = \begin{cases} \frac{\pi(\mathcal{I})}{\Pi([\phi])}, & \text{if } \mathcal{I} \models \phi \text{ and } \mathcal{B}(\neg\phi) < 1; \\ 1, & \text{if } \mathcal{I} \models \phi \text{ and } \mathcal{B}(\neg\phi) = 1; \\ \min\{\pi(\mathcal{I}), (1-\tau)\}, & \text{if } \mathcal{I} \not\models \phi. \end{cases} \tag{13}$$

The condition $\mathcal{B}(\neg\phi) < 1$ in Equation 13 is equivalent to $\exists \mathcal{I}' : \mathcal{I}' \models \phi \Rightarrow \pi(\mathcal{I}') > 0$, i.e., $\Pi([\phi]) > 0$; likewise, the condition $\mathcal{B}(\neg\phi) = 1$ is equivalent to $\Pi([\phi]) = 0$, which implies $\pi(\mathcal{I}) = 0$ $\forall \mathcal{I} \models \phi$. Therefore, the second case in Equation 13 provides for the *revision* of beliefs that are in contradiction with new information ϕ. We can notice that in general, the operator treats new information ϕ in the negative sense: being told ϕ denies the possibility of world situations where ϕ is false (third case of Equation 13). The possibility of world situations where ϕ is true may only increase due to the first case in equation 13 or revision (second case of Equation 13). We can notice that in [18], if a sure information contradicts an existing proposition that is fully believed then it leads to inconsistency. Revising with our operator instead, leads the agent to believe the more recent information and give up the oldest to restore consistency. The reason is that we suppose that the agent is not in a purely static context. Therefore, as convincingly pointed out by Delgrande and colleagues in [17], there is reason for giving priority to more recent items.

In the following, we show that the above belief change operator obeys the AGM belief revision rationality postulates. After recalling that the *expansion* of a crisp set of formulas K with a formula $\phi \in \mathcal{L}$ is $K + \phi = \{\psi : K \cup \{\phi\} \vdash \psi\}$, let us define the expansion of a fuzzy set of formulas $\mathcal{B}$ with a formula $\phi \in \mathcal{L}$ from a source trusted to degree τ, for all $\psi \in \mathcal{L}$, as

$$\left(\mathcal{B} + \frac{\tau}{\phi}\right)(\psi) = \max\{\alpha : (\mathcal{B} \cup \{(\phi, \tau)\})_\alpha \vdash \psi\}, \tag{14}$$

where $(\mathcal{B} \cup \{(\phi, \tau)\})_\alpha$ is the crisp set corresponding to the α-cut of the fuzzy set $\mathcal{B} \cup \{(\phi, \tau)\}$. In terms of possibility distribution, this corresponds to

$$\left(\pi + \frac{\tau}{\phi}\right)(\mathcal{I}) = \min\{\pi(\mathcal{I}), \phi^{\mathcal{I}} + (1 - \phi^{\mathcal{I}})(1 - \tau)\}. \tag{15}$$

The AGM revision rationality postulates K∗1–K∗8 [21] may be reformulated as follows in a possibilistic setting (with some slight but important differences from [18], which mostly have to do with the fact that we deal here with partially trusted new information), where π is a normalized possibility distribution inducing $\mathcal{B}$, $\phi \in \mathcal{L}$ is a formula and $\pi' = \pi * \frac{\tau}{\phi}$ is the possibility distribution inducing $\mathcal{B}' = \mathcal{B} * \frac{\tau}{\phi}$:

$\mathcal{B}*1$ π' is a normalized possibility distribution.

$\mathcal{B}*2$ $\mathcal{B}'(\phi) \geq \tau$ (priority to new information).

$\mathcal{B}*3$ $\mathcal{B}' \subseteq \mathcal{B} + \frac{\tau}{\phi}$, or, equivalently, for all $\mathcal{I} \in \Omega$, $\pi'(\mathcal{I}) \geq \min\{\pi(\mathcal{I}), \phi^{\mathcal{I}} + (1 - \phi^{\mathcal{I}})(1 - \tau)\}$, i.e., revising does not yield more specific results than expanding.

$\mathcal{B}*4$ If $\mathcal{B}(\neg\phi) = 0$, then $\mathcal{B} + \frac{\tau}{\phi} \subseteq \mathcal{B}'$, or, equivalently, for all $\mathcal{I} \in \Omega$, $\pi'(\mathcal{I}) \leq \min\{\pi(\mathcal{I}), \phi^{\mathcal{I}} + (1 - \phi^{\mathcal{I}})(1 - \tau)\}$, i.e., if ϕ is not rejected by π, revision reduces to expansion;

$\mathcal{B}*5$ $\pi'(\mathcal{I}) \leq 1 - \tau$ for all $\mathcal{I} \in \Omega$ if and only if $\phi \equiv \bot$.

$\mathcal{B}*6$ If $\phi \equiv \psi$, then $\mathcal{B} * \frac{\tau}{\phi} = \mathcal{B} * \frac{\tau}{\psi}$.

$\mathcal{B}*7$ $\mathcal{B} * \frac{\tau}{\phi \wedge \psi} \subseteq \mathcal{B}' + \frac{\tau}{\psi}$, or, equivalently, for all $\mathcal{I} \in \Omega$, $(\pi * \frac{\tau}{\phi \wedge \psi})(\mathcal{I}) \geq \min\{\pi'(\mathcal{I}), \psi^{\mathcal{I}} + (1 - \psi^{\mathcal{I}})(1 - \tau)\}$.

$\mathcal{B}*8$ If $\Delta([\phi \wedge \neg\psi]) \geq 1 - \tau$, then, $\mathcal{B}' + \frac{\tau}{\psi} \subseteq \mathcal{B} * \frac{\tau}{\phi \wedge \psi}$, or, equivalently, for all $\mathcal{I} \in \Omega$, $(\pi * \frac{\tau}{\phi \wedge \psi})(\mathcal{I}) \leq \min\{\pi'(\mathcal{I}), \psi^{\mathcal{I}} + (1 - \psi^{\mathcal{I}})(1 - \tau)\}$.

PROPOSITION 1. *For all $\phi \in \mathcal{L}$, the belief change operator $*$ of Definition 8 obeys postulates $\mathcal{B}*1$–$\mathcal{B}*8$.*

Proof: $\mathcal{B}*1$ holds because by definition of π' (see equation 13): (i) if $\Pi([\phi]) > 0$ then $\exists \mathcal{I}_0 \models \phi$ such that $\pi'(\mathcal{I}_0) = 1$, $(\pi(\mathcal{I}_0) = \Pi([\phi]))$; and (ii) if $\Pi([\phi]) = 0$ then $\forall \mathcal{I} \models \phi$, $\pi'(\mathcal{I}) = 1$.

As for $\mathcal{B}*2$,

$$\begin{aligned} \mathcal{B}'(\phi) &= 1 - \max_{\mathcal{I} \not\models \phi} \pi'(\mathcal{I}) \\ &= 1 - \max_{\mathcal{I} \not\models \phi} \min\{\pi(\mathcal{I}), 1 - \tau\} \\ &\geq 1 - (1 - \tau) = \tau. \end{aligned}$$

To prove $\mathcal{B}*3$, we consider the three possible cases:

(i) $\phi^{\mathcal{I}} = 1$ and $\mathcal{B}(\neg\phi) = 1$: $\pi'(\mathcal{I}) = 1 \geq \min\{\pi(\mathcal{I}), \phi^{\mathcal{I}} + (1 - \phi^{\mathcal{I}})(1 - \tau)\}$;

(ii) $\phi^{\mathcal{I}} = 1$ and $\mathcal{B}(\neg\phi) < 1$: $\pi'(\mathcal{I}) = \frac{\pi(\mathcal{I})}{\Pi([\phi])} \geq \pi(\mathcal{I})$;

(iii) $\phi^{\mathcal{I}} = 0$: $\min\{\pi(\mathcal{I}), 1 - \tau\} = \min\{\pi(\mathcal{I}), 1 - \tau\}$.

The proof of $\mathcal{B}*4$ is similar to the one of $\mathcal{B}*3$. The only problematic case would be when $\Pi([\phi]) = 0$. However, this case is impossible thanks to the hypothesis $\mathcal{B}(\neg\phi) = 0$, which implies $\Pi([\phi]) = 1$.

The proof of $\mathcal{B}*5$ can be done in two steps. To begin with, we have to prove that if $[\phi] \neq \emptyset$ then $\exists \mathcal{I}_0 \in \Omega$ such that $\pi'(\mathcal{I}_0) > 1 - \tau$. If $\Pi([\phi]) = 0$ then $\pi'(\mathcal{I}) = 1$, $\forall \mathcal{I}$. Otherwise, if $\Pi([\phi]) > 0$, it is enough to consider the $\mathcal{I}_0$ which is such that $\pi(\mathcal{I}_0) = \Pi([\phi])$.

Secondly, we have to prove that, if $[\phi] = \emptyset$, then $\forall \mathcal{I} \in \Omega\ \pi'(\mathcal{I}) \leq 1 - \tau$. Now, $[\phi] = \emptyset$ means that $\not\exists \mathcal{I} \in \Omega$ such that $\mathcal{I} \models \phi$. This means that $\forall \mathcal{I} \in \Omega,\ \mathcal{I} \not\models \phi$ and then $\forall \mathcal{I} \in \Omega$ we have $\pi'(\mathcal{I}) = \min\{\pi(\mathcal{I}), 1 - \tau\} \leq 1 - \tau$.

Let us consider that π'_ϵ corresponds to the resulting possibility distribution after the acquisition of new information ϵ. Proving $\mathcal{B}{*}6$ amounts to proving that, if $\phi \equiv \psi$, then $\forall \mathcal{I} \in \Omega,\ \pi'_\phi(\mathcal{I}) = \pi'_\psi(\mathcal{I})$. Now, $\phi \equiv \psi$ means that $[\phi] = [\psi]$ and then $\Pi([\phi]) = \Pi([\psi])$. A direct consequence is that $\pi'_\phi(\mathcal{I}) = \pi'_\psi(\mathcal{I})$.

The proofs of $\mathcal{B}{*}7$ and $\mathcal{B}{*}8$ are rather lengthy, having to be broken up into several cases; therefore, they will be omitted due to lack of space. □

5. DESIRES

We suppose that the agent's subjective qualitative utilities are determined dynamically through a rule-based deliberation mechanism. Associating a qualitative utility first to worlds and not to formulas allows us to (i) directly construct the possibility distribution u; and (ii) makes it possible to also calculate the qualitative degree of formulas which do not appear explicitly on the right-hand side of any rule.

Like in [4], the qualitative utility associated to each positive desire formula is computed on the basis of the guaranteed possibility measure Δ.

The set of the agent's justified positive desires, $\mathcal{J}$, is induced by the assignment of a qualitative utility u, which, unlike π, needs not be normalized, since desires may very well be inconsistent.

DEFINITION 9 (JUSTIFIED DESIRE). *Given a qualitative utility assignment u (formally a possibility distribution), the degree to which the agent desires $\phi \in \mathcal{L}$ is given by*

$$\mathcal{J}(\phi) = \Delta([\phi]) = \min_{\mathcal{I} \models \phi} u(\mathcal{I}). \tag{16}$$

In words, the degree of justification of a desire is given by the guaranteed qualitative utility of the set of all worlds in which the desire would be fulfilled. Intuitively, a desire is justified to the extent that all the worlds in which it is fulfilled are desirable.

Interpreting $\mathcal{J}(\phi)$ as a degree of membership defines the fuzzy set $\mathcal{J}$ of the agent's justified positive desires.

In turn, a qualitative utility assignment u is univocally determined by the mental state of the agent as explained below.

DEFINITION 10 (RULE ACTIVATION). *Let $R = \beta_R, \psi_R \Rightarrow^+_D \phi$ be a desire-generation rule. The degree af activation of R,* $\mathrm{Deg}(R)$*, is given by*

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

For an unconditional rule $R = \alpha_R \Rightarrow^+_D \phi$,

$$\mathrm{Deg}(R) = \alpha_R.$$

REMARK 1. *Here, we assume the commensurability between belief degrees and desire degrees in order to make a direct comparison possible between belief and desire degrees.*

Let us denote by $\mathcal{R}^{\mathcal{I}}_J = \{R \in \mathcal{R}_J : \mathcal{I} \models \mathrm{rhs}(R)\}$ the subset of $\mathcal{R}_J$ containing just the rules whose right-hand side would be true in world $\mathcal{I}$.

DEFINITION 11 (DESIRED WORLDS). *The qualitative utility assignment $u : \Omega \to [0, 1]$ (formally a possibility distribution) is defined, for all $\mathcal{I} \in \Omega$, as*

$$u(\mathcal{I}) = \max_{R \in \mathcal{R}^{\mathcal{I}}_J} \mathrm{Deg}(R). \tag{17}$$

At first glance, Definitions 9, 10, and 11 appear to be circular. After all, u depends on the degree of activation of some rules, which in turn depends on the degree of justification of some desires, which in turn depends on u! However, this apparent circularity may be resolved by an algorithmic translation of the two definitions which reveal u is in fact the limit distribution obtained by iteratively applying the two definitions.

Given a mental state $\mathcal{S} = \langle \pi, \mathcal{R}_J \rangle$, the following algorithm computes the corresponding qualitative utility assignment, u.

ALGORITHM 1 (DELIBERATION).

1. $i \leftarrow 0$*; for all* $\mathcal{I} \in \Omega$, $u_0(\mathcal{I}) \leftarrow 0$*;*
2. $i \leftarrow i + 1$*;*
3. *For all* $\mathcal{I} \in \Omega$,
$$u_i(\mathcal{I}) \leftarrow \begin{cases} \max_{R \in \mathcal{R}^{\mathcal{I}}_J} \mathrm{Deg}_{i-1}(R), & \text{if } \mathcal{R}^{\mathcal{I}}_J \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$
where $\mathrm{Deg}_{i-1}(R)$ *is the degree of activation of rule R calculated using u_{i-1} as the qualitative utility assignment;*
4. *if* $\max_{\mathcal{I}} |u_i(\mathcal{I}) - u_{i-1}(\mathcal{I})| > 0$*, i.e., if a fixpoint has not been reached yet, go back to Step 2;*
5. *For all* $\mathcal{I} \in \Omega$, $u(\mathcal{I}) \leftarrow u_i(\mathcal{I})$*; u is the qualitative utility assignment corrisponding to mental state $\mathcal{S}$.*

Let $\mathrm{Img}(u) = \{\alpha : \exists \mathcal{I}\ u(\mathcal{I}) = \alpha\}$ be the image of u.

PROPOSITION 2. $\|\mathrm{Img}(u)\| \leq \|\mathcal{R}_J\| + 1$.

Proof: We will prove the thesis by constructing a finite level set Λ such that $\mathrm{Img}(u) \subseteq \Lambda$.

According to Definition 11, $u(\mathcal{I})$ can only take values $\mathrm{Deg}(R)$, for some rule $R \in \mathcal{R}_J$, or 0. For all unconditional rules R, let $\alpha_R \in \Lambda$. For all conditional rules R of the form $\beta_R, \psi_R \Rightarrow^+_D \phi$, in a given mental state, $\mathcal{B}(\beta_R) = b_R$ is constant; let $b_R \in \Lambda$. Of course, $\mathrm{Deg}(R) = \min\{b_R, \mathcal{J}(\psi_R)\}$, but, by Definition 9, $\mathcal{J}(\phi) = \min_{\mathcal{I} \models \psi_R} u(\mathcal{I}) \in \Lambda$, for min just selects one of its argument and cannot create new values.

Now, by construction, $\|\Lambda\| \leq \|\mathcal{R}_J\| + 1$, since $0 \in \Lambda$ and there is at most one distinct α_R for each unconditional rule and one distinct b_R for each conditional rule. Λ contains all values $u(\mathcal{I})$ might conceivably take up, although the actual values may be fewer if, for some R, it turns out that $\mathcal{J}(\psi_R) < b_R$. Therefore $\mathrm{Img}(u) \subseteq \Lambda$ and $\|\mathrm{Img}(u)\| \leq \|\Lambda\| \leq \|\mathcal{R}_J\| + 1$. □

PROPOSITION 3. *Algorithm 1 always terminates.*

Proof: First of all, we will prove, by induction, that, at each iteration $i > 0$, for all $\mathcal{I}$, $u_i(\mathcal{I}) \geq u_{i-1}(\mathcal{I})$.

This is certainly true of $i = 1$, for $u_0(\mathcal{I}) = 0$ for all $\mathcal{I}$ and $u_1(\mathcal{I}) \geq u_0(\mathcal{I}) = 0$.

Now, for all iteration $i > 1$, we assume $u_{i-1}(\mathcal{I}) \geq u_{i-2}(\mathcal{I})$ for all $\mathcal{I}$ and will prove that $u_i(\mathcal{I}) \geq u_{i-1}(\mathcal{I})$.

We need a few definitions. First of all, let us partition $\mathcal{R}_J$ into the set U_J of unconditional rules and the set C_J of conditional rules. Now, given $\mathcal{I}$, let $U^{\mathcal{I}}_J$ be the set of unconditional rules R such that $\mathcal{I} \models \mathrm{rhs}(R)$ and $C^{\mathcal{I}}_J$ be the set of conditional rules R such that $\mathcal{I} \models \mathrm{rhs}(R)$. Let $\alpha_{\mathcal{I}} = \max\{\alpha_R : R \in U^{\mathcal{I}}_J\}$, $\alpha_{\mathcal{I}} = 0$ if $U^{\mathcal{I}}_J = \emptyset$. Now we can write

$$u_i(\mathcal{I}) = \max\{\alpha_{\mathcal{I}}, \max_{R \in C^{\mathcal{I}}_J} \min\{b_R, \min_{\mathcal{I}' \models \psi_R} u_{i-1}(\mathcal{I}')\}\}. \tag{18}$$

There are two cases:

1. $u_{i-1}(\mathcal{I}) \leq \alpha_{\mathcal{I}}$: in this case, $u_i(\mathcal{I}) \geq \alpha_{\mathcal{I}} \geq u_{i-1}(\mathcal{I})$, and the inductive thesis holds;

2. $u_{i-1}(\mathcal{I}) > \alpha_{\mathcal{I}}$: in this case,
$$u_i(\mathcal{I}) = \max_{R \in C_J^{\mathcal{I}}} \min\{b_R, \min_{\mathcal{I}' \models \psi_R} u_{i-1}(\mathcal{I}')\},$$
and $u_i(\mathcal{I}) \geq u_{i-1}(\mathcal{I})$ may be rewritten as
$$\begin{aligned} \max\nolimits_{R \in C_J^{\mathcal{I}}} \min\{b_R, \min\nolimits_{\mathcal{I}' \models \psi_R} u_{i-1}(\mathcal{I}')\} &\geq \\ \max\nolimits_{R \in C_J^{\mathcal{I}}} \min\{b_R, \min\nolimits_{\mathcal{I}' \models \psi_R} u_{i-2}(\mathcal{I}')\}, & \end{aligned}$$
which will certainly hold if, for all $R \in \mathcal{R}_J$,
$$\min_{\mathcal{I}' \models \psi_R} u_{i-1}(\mathcal{I}') \geq \min_{\mathcal{I}' \models \psi_R} u_{i-2}(\mathcal{I}');$$
but this must be true, since, by the inductive hypothesis, for all $\mathcal{I}'$, $u_{i-1}(\mathcal{I}') \geq u_{i-2}(\mathcal{I}')$.

This proves that, for all $i > 0$ and for all $\mathcal{I}$, $u_i(\mathcal{I}) \geq u_{i-1}(\mathcal{I})$.

Now, since, by Proposition 2, $\mathrm{Img}(u)$ is finite, after a finite number of iterations, $u_i(\mathcal{I}) = u_{i-1}(\mathcal{I})$ for all $\mathcal{I}$, and the algorithm will terminate. □

The following example is attributed by Derek Baker [1] to John Williams. Bob desires to date Sally, and desires to date Sue, without desiring that he dates both. This may be translated into our formalism by the following rule:
$$1 \Rightarrow_D^+ (\mathrm{sue} \vee \mathrm{sally}) \wedge \neg(\mathrm{sue} \wedge \mathrm{sally}).$$
For the sake of simplicity, let us say the set of interpretations is
$$\Omega = \left\{ \begin{array}{lcl} \mathcal{I}_0 & = & \{\mathrm{sally} \mapsto 0, \mathrm{sue} \mapsto 0\}, \\ \mathcal{I}_1 & = & \{\mathrm{sally} \mapsto 0, \mathrm{sue} \mapsto 1\}, \\ \mathcal{I}_2 & = & \{\mathrm{sally} \mapsto 1, \mathrm{sue} \mapsto 0\}, \\ \mathcal{I}_3 & = & \{\mathrm{sally} \mapsto 1, \mathrm{sue} \mapsto 1\} \end{array} \right\}.$$
By applying the deliberation algorithm, the generated qualitative utility assignment is
$$u(\mathcal{I}_0) = 0, \quad u(\mathcal{I}_1) = 1, \quad u(\mathcal{I}_2) = 1, \quad u(\mathcal{I}_3) = 0.$$
From here, we can compute the degree of justification of various of Bob's hypothetical desires; for instance, we will discover that
$$\begin{aligned} \mathcal{J}(\mathrm{sue}) &= \min_{\mathcal{I} \models \mathrm{sue}} u(\mathcal{I}) = \min\{u(\mathcal{I}_1), u(\mathcal{I}_3)\} = 0; \\ \mathcal{J}(\mathrm{sally}) &= \min_{\mathcal{I} \models \mathrm{sally}} u(\mathcal{I}) = \min\{u(\mathcal{I}_2), u(\mathcal{I}_3)\} = 0; \\ \mathcal{J}(\mathrm{sally} \wedge \neg\mathrm{sue}) &= \min_{\mathcal{I} \models \mathrm{sally} \wedge \neg\mathrm{sue}} u(\mathcal{I}) = u(\mathcal{I}_2) = 1; \\ \mathcal{J}(\mathrm{sue} \wedge \neg\mathrm{sally}) &= \min_{\mathcal{I} \models \mathrm{sue} \wedge \neg\mathrm{sally}} u(\mathcal{I}) = u(\mathcal{I}_1) = 1. \end{aligned}$$

This is a prototypical example of an inconsistent set of justfied desires.

6. GOALS

Here, we make a clear distinction between desires and goals. As pointed out in the previous sections, we suppose that desires may be inconsistent. Goals, instead, are defined as a consistent subset of desires.

DEFINITION 12. *The overall possibility of a set $S \subseteq \mathcal{L}$ of formulas is*
$$\Pi([S]) = \max_{\mathcal{I} \in [S]} \pi(\mathcal{I}). \tag{19}$$

The following definition extends $\mathcal{J}$, the degree of justification of a desire, to sets of desires.

DEFINITION 13. *The overall qualitative utility, or justification, of a set $S \subseteq \mathcal{L}$ of formulas is*
$$\mathcal{J}(S) = \Delta([S]) = \min_{\mathcal{I} \in [S]} u(\mathcal{I}). \tag{20}$$

It follows from the properties of the minumum guaranteed possibility, that
$$\mathcal{J}(S) = \Delta([S]) = \Delta\left(\bigcap_{\phi \in S} [\phi]\right) \geq \max_{\phi \in S}\{\Delta([\phi])\} = \max_{\phi \in S}\{\mathcal{J}(\phi)\}. \tag{21}$$

Therefore, the addition of a desire to a set of desires can only lead to an increase of the justification level of the resulting enlarged set of desires.

PROPOSITION 4. *Let $S \subseteq \mathcal{L}$ be a set of desires. For all desire ϕ,*
$$\mathcal{J}(S \cup \{\phi\}) \geq \mathcal{J}(S); \tag{22}$$
$$\mathcal{J}(S) \geq \mathcal{J}(S \setminus \{\phi\}). \tag{23}$$

Proof: By Definition 4, $[S \cup \{\phi\}] = [S] \cap [\phi]$. Therefore, by the properties of the minimum guaranteed possibility, we can write
$$\begin{aligned} \mathcal{J}(S \cup \{\phi\}) = \Delta([S] \cap [\phi]) &\geq \max\{\Delta([S]), \Delta([\phi])\} \\ &\geq \Delta([S]) = \mathcal{J}(S). \end{aligned}$$
The proof of Equation 23 is obtainend by replacing S with $S' \setminus \{\phi\}$ in Equation 22. □

This fits very nicely with the intuition of the man in the street that, e.g., if Sally likes the idea of marrying a rich man and she also like the idea of marrying a handsome man, all the more she will like the idea of marrying a rich, handsome man. Say Sally's desire-generation rules are
$$\begin{array}{lcl} 0.7 & \Rightarrow_D^+ & \mathrm{rich}, \\ 0.8 & \Rightarrow_D^+ & \mathrm{handsome}. \end{array}$$

Applying these rules yields $\mathcal{J}(\{\mathrm{rich}\}) = 0.7$, $\mathcal{J}(\{\mathrm{handsome}\}) = 0.8$, and $\mathcal{J}(\{\mathrm{rich}, \mathrm{handsome}\}) = 0.8$.

A rational agent will select as goals the set of desires that, besides being logically "consistent", is also maximally desirable, i.e., maximally justified. The problem with logical "consistency", however, is that it does not capture "implicit" inconsistencies among desires, that is consistency due to the agent beliefs (I adopt as goals only desires which are not inconsistent with my beliefs). Therefore, a suitable definition of desire consistency in the possibilistic setting is required. Such definition must take the agent's cognitive state into account as pointed out, for example, in [1, 11, 25].

For example, an agent desires p and desires q, believing that $p \supset \neg q$. Although $\{p, q\}$, as a set of formulas, i.e., syntactically, is logically consistent, it is not if one take the belief $p \supset \neg q$ into account.

We argue that a suitable definition of such "cognitive" consistency is one based on the possibility of the set of desires, as defined above. Indeed, a set of desires S is consistent, in the cognitive sense, if and only if $\Pi([S]) > 0$. Of course, cognitive consistency implies logical consistency: if S is logically inconsistent, $\Pi([S]) = 0$. We will take a step forward, by assuming a rational agent will select as goals the most desirable set of desires among the most possible such sets.

Let $\mathcal{D} = \{S \subseteq \text{supp}(\mathcal{J})\}$, i.e., the set of desire sets whose justification is greater than zero.

DEFINITION 14. *Given $\gamma \in (0, 1]$,*

$$\mathcal{D}_\gamma = \{S \in \mathcal{D} : \Pi([S]) \geq \gamma\}$$

is the subset of $\mathcal{D}$ containing only those sets whose overall possibility is at least γ.

For every given level of possibility γ, a rational agent will elect as its goal set the maximally desirable of the γ-possible sets.

DEFINITION 15 (GOAL SET). *The γ-possible goal set is*

$$G_\gamma = \begin{cases} \arg\max_{S \in \mathcal{D}_\gamma} \mathcal{J}(S) & \textit{if } \mathcal{D}_\gamma \neq \emptyset, \\ \emptyset & \textit{otherwise.} \end{cases}$$

We denote by γ^* the maximum possibility level such that $G_\gamma \neq \emptyset$. Then, the goal set elected by a rational agent will be

$$G^* = G_{\gamma^*}, \quad \gamma^* = \max_{G_\gamma \neq \emptyset} \gamma. \tag{24}$$

7. CONCLUSION

A theoretical framework for goal generation in BDI agent has been justified and developed. Beliefs and desires are represented by means of two possibility distributions. A deliberative process is responsible for generating the distribution of qualitative utility that underlies desire justification, and the election of goals considers their cognitive consistency, realized as possibility.

A limit of the framework as described in this paper is that it does not take negative preferences (which are, in a sense, the other face of desires) into account. This limit is not inherent in any of the choices we have made, and we plan on including an account of negative preferences in the future.

8. REFERENCES

[1] D. Baker. Ambivalent desires and the problem with reduction. *Philosophical Studies*, Published online: 25 March, 2009.

[2] J. Bell and Z. Huang. Dynamic goal hierarchies. In *PRICAI '96*, pages 88–103. Springer-Verlag, 1997.

[3] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar possibility theory in preference modeling: Representation, fusion and optimal solutions. *Inf. Fusion*, 7(1):135–150, 2006.

[4] S. Benferhat and S. Kaci. Logical representation and fusion of prioritized information based on guaranteed possibility measures: application to the distance-based merging of classical bases. *Artif. Intell.*, 148(1-2):291–333, 2003.

[5] J. Blee, D. Billington, and A. Sattar. Reasoning with levels of modalities in BDI logic. In *PRIMA 2007*, pages 410–415. Springer-Verlag, 2009.

[6] C. Boutilier. Toward a logic for qualitative decision theory. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning*, pages 75–86, 1994.

[7] 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.

[8] J. T. Cacioppo and G. G. Berntson. The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science*, 8:133–137, 1999.

[9] A. Casali, L. Godo, and C. Sierra. Graded BDI models for agent architectures. In *Computational logic in multiagent systems. Fifth international workshop, CLIMA V. Pre-proceedings*, pages 18–33, 2004.

[10] C. Castelfranchi and F. Paglieri. The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions. *Synthese*, 155(2):237–263, 2007.

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

[12] C. da Costa Pereira and A. Tettamanzi. Goal generation and adoption from partially trusted beliefs. In *Proceedings of ECAI 2008*, pages 453–457. IOS Press, 2008.

[13] C. da Costa Pereira and A. Tettamanzi. Goal generation with relevant and trusted beliefs. In *Proceedings of AAMAS'08*, pages 397–404. IFAAMAS, 2008.

[14] R. da Silva Neves and E. Raufaste. A psychological study of bipolarity in the possibilistic framework. In *IPMU'2004*, pages 975–981, 2004.

[15] M. Dastani, J. Hulstijn, and L. van der Torre. Bdi and qdt: a comparison based on classical decision theory. In *In Proceedings of AAAI Spring Symposium GTDT'01*, 2001.

[16] B. De Baets, E. Tsiporkova, and R. Mesiar. Conditioning in possibility theory with strict order norms. *Fuzzy Sets Syst.*, 106(2):221–229, 1999.

[17] J. Delgrande, D. Dubois, and J. Lang. Iterated revision as prioritized merging. In *KR*, pages 210–220, 2006.

[18] D. Dubois and H. Prade. A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. *International Journal of Approximate Reasoning*, 17:295–324, 1997.

[19] D. Dubois and H. Prade. Possibility theory, probability theory and multiple-valued logics: A clarification. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):35–66, 2001.

[20] D. Dubois and H. Prade. An overview of the asymmetric bipolar representation of positive and negative information in possibility theory. *Fuzzy Sets Syst.*, 160(10):1355–1366, 2009.

[21] P. Gärdenfors. Belief revision: A vademecum. In *Meta-Programming in Logic*, pages 1–10. Springer, Berlin, 1992.

[22] J. Lang. Conditional desires and utilities: an alternative logical approach to qualitative decision theory. In *ECAI*, pages 318–322, 1996.

[23] J. Lang, L. Van Der Torre, and E. Weydert. Utilitarian desires. *Autonomous Agents and Multi-Agent Systems*, 5(3):329–363, 2002.

[24] S. Parsons and P. Giorgini. An approach to using degrees of belief in BDI agents. In *Information, Uncertainty, Fusion*. Kluwer, 1999.

[25] A.S. Rao and M.P. Georgeff. Asymmetry thesis and side-effect problems in linear-time and branching-time intention logics. In *IJCAI*, pages 498–505, 1991.

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

[27] S. Tan and J. Pearl. Qualitative decision theory. In *AAAI*, pages 928–933, 1994.

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

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