ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-606-5-641

Belief-Goal Relationships in Possibilistic Goal Generation

Célia da Costa Pereira and **Andrea G. B. Tettamanzi** [1]

Abstract.

The way in which the relationships between beliefs, goals, and intentions are captured by a formalism can have a significant impact on the design of a rational agent. In particular, what Rao and Georgeff underline about the relationships between goals and beliefs is that it is reasonable to require a rational agent not to allow goal-belief inconsistency, while goal-belief incompleteness can be allowed.

We study a theoretical framework, grounded in possibility theory, which (i) accounts for the aspects involved in representing and changing beliefs and goals, and (ii) obeys Rao and Georgeff's requirement. We propose a formalization of a possibilistic extension of Bratman's *asymmetry thesis* to hold between goals and beliefs. Finally, we show that our formalism avoids the *side-effect* and the *transference* problems.

1 Introduction

The need of adapting the traditional BDI (Beliefs, Desires and Intentions) agent model, based on the concept of practical reasoning [4], to make it more suitable to represent (i) the changing world properties, (ii) the fact that available information is often partial, and (iii) the fact that uncertainty is omnipresent in realistic domains, is not new. Much work extending the BDI model has been proposed. Parsons and Giorgini [14] proposed to treat beliefs as degrees of evidence. They developed a system based upon an existing formal multi-context-based model of the agent's mental attributes in order to permit beliefs, desires, and intentions to admit degrees. The multi-context component is also adopted by Casali and colleagues [5], who proposed a general model for graded BDI agents and an architecture able to model the agent's graded mental attitudes. Like Casali, Blee and colleagues [3] introduce levels in all the mentalistic notions of BDI, as well as using numeric, possibilistic-type functions in its semantics. The difference between the two is that Blee's framework, unlike Casali's, uses a common syntax for the agent's mental atittudes. All these proposals are only focused on the formal aspects of the representation of the graded mental attitudes. How these degrees arise and the fact that desire/goal degrees are not always independent of the degrees of beliefs are not considered.

We study the properties of a general possibilistic framework recently proposed by us [8]. That formalism considers the cognitive relationships between beliefs and goals [6] and represents beliefs and desires as possibility distributions following [11]. In addition, the agent's desires and its utility degrees are generated by means of rules.

The asymmetry thesis was originaly formulated by Bratman [4], who stated that, in a formalism for representing *rational agents*, beliefs and actions should be asymmetrical with respect to direct voluntary control. Rao and Georgeff extended the thesis to hold in belief-goal and in goal-intention relationships. We show that the formalism under study obeys the possibilistic extension of Rao and Georgeff's *Asymmetry Thesis* for belief-goal relationships. Finally, we also prove that the formalism avoids the *side-effect* and the *transference* problems.

[1] Università degli Studi di Milano, Italy, email: {celia.pereira,andrea.tettamanzi}@unimi.it

2 Possibilistic Representation

In this section, we review the possibilistic representation of beliefs and desires proposed in [8].

2.1 Possibility Theory

Possibility theory is a mathematical theory of uncertainty that relies upon fuzzy set theory [19], in that the (fuzzy) set of possible values for a variable of interest is used to describe the uncertainty as to its precise value. The membership function of such set, π, is called a *possibility distribution* and its range is $[0, 1]$.

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:*

$$\Pi(A) = \max_{s \in A} \pi(s); \quad (1)$$

$$N(A) = 1 - \Pi(\bar{A}) = \min_{s \in \bar{A}} \{1 - \pi(s)\}. \quad (2)$$

Another interesting measure that can be defined based on a possibility distribution is *guaranteed possibility* [11].

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

$$\Delta(A) = \min_{s \in A} \pi(s); \quad (3)$$

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. $\Pi(A \cap B) \leq \min\{\Pi(A), \Pi(B)\}$;
3. $\Pi(\emptyset) = N(\emptyset) = 0; \quad \Pi(\Omega) = N(\Omega) = 1$;
4. $N(A \cap B) = \min\{N(A), N(B)\}$;
5. $N(A \cup B) \geq \max\{N(A), N(B)\}$;

6. $\Pi(A) = 1 - N(\bar{A})$ (duality);
7. $N(A) > 0 \Rightarrow \Pi(A) = 1; \quad \Pi(A) < 1 \Rightarrow N(A) = 0;$
8. $\Delta(A) \leq \Pi(A); \quad N(A) \leq \Pi(A);$
9. $\Delta(A \cup B) = \min\{\Delta(A), \Delta(B)\};$
10. $\Delta(A \cap B) \geq \max\{\Delta(A), \Delta(B)\}.$

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

2.2 Language and Interpretations

A classical propositional language may be used to represent information for manipulation by a cognitive agent.

Definition 3 (Language) *Let $\mathcal{A}$ be a* finite[2] *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 *$[\phi]$ will denote the set of all models of a formula $\phi \in \mathcal{L}$, $[\phi] = \{\mathcal{I} \in \Omega : \mathcal{I} \models \phi\}$. By extension, if $S \subseteq \mathcal{L}$ is a set of formulas, $[S] = \{\mathcal{I} \in \Omega : \forall \phi \in S, \mathcal{I} \models \phi\} = \bigcap_{\phi \in S} [\phi]$.*

2.3 Representing Beliefs and Desires

Beliefs and desires of a cognitive agent describe situations in the past, present, and future and are represented thanks to two different possibility distributions π and u respectively. While π is normalized because an agent's beliefs are supposed to be consistent, u does not, since desires may very well be inconsistent.

A description of how desires arise is given in terms of desire-generation rules, which are a possibilistic extension of van Riemsdijk's "desire-adoption" rules [18].

Definition 5 (Desire-Generation Rule) *A desire-generation rule R is an expression of the form $\beta_R, \psi_R \Rightarrow_D^+ \phi$, where $\beta_R, \psi_R, \phi \in \mathcal{L}$. An unconditional desire-generation rule has the form $\alpha \Rightarrow_D^+ \phi$, 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 with $\mathrm{rhs}(R)$ the formula on the right-hand side of R.

A *belief* is regarded as a necessity degree induced by a normalized possibility distribution

$$\pi : \Omega \to [0,1], \tag{4}$$

which represents a plausibility order of possible states of affairs: $\pi(\mathcal{I})$ is the possibility degree of interpretation $\mathcal{I}$.

[2] Like in [2], 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 [9].

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:*

$$\mathcal{B}(\phi) = N([\phi]) = 1 - \max_{\mathcal{I} \not\models \phi} \{\pi(\mathcal{I})\}. \tag{5}$$

2.4 Mental State

The mental state of an agent consists of its beliefs and the rules that define 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.*

Example Dr. A. Gent has submitted a paper to ECAI 2010 he has written with his co-author I. M. Flaky, who has promised to go to Lisbon to present it if it is accepted. Dr. Gent knows that, if the paper is accepted, publishing it (which is his great desire), means to pay the conference registration (for his co-author or for himself) and then be ready to go to Lisbon to present it, in case I. M. is unavailable.

If the paper is accepted (a), Dr. Gent is willing to pay the registration (r); furthermore, if the paper is accepted and Dr. Flaky turns out to be unavailable (q), he is willing to go to Lisbon to present it (p). Finally, if he knows the paper is accepted and wishes to present it, he will desire to have a hotel room (h) and a plane ticket reserved (t).

Dr. Gent has some *a priori* beliefs about this situation, namely that if the hotels are full (f), he will not succeed in booking a hotel room; similarly, he believes that if the planes are all booked out (b), he will not succeed in reserving a flight, although this is not necessarily true, if he puts himself in the waiting list and a reservation is cancelled. Finally, he believes the organizers will enforce the rule whereby his paper will be presented only if it is accepted and a registration is paid.

The set of atomic propositions is then $\mathcal{A} = \{a, b, f, h, p, r, t, q\}$.

Dr. Gent's $\mathcal{R}_J$ may be described by the following desire-generation rules:

$$\begin{array}{lrcl} R_1: & a, & p & \Rightarrow_D^+ \quad t \wedge h, \\ R_2: & a \wedge q, & \top & \Rightarrow_D^+ \quad p, \\ R_3: & a, & \top & \Rightarrow_D^+ \quad r. \end{array}$$

2.5 Updating Beliefs

Agents update their possibility distribution π in light of new information $\phi \in \mathcal{L}$ coming from a source trusted to a certain extent $\tau \in [0,1]$ by means of the following belief change operator.

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 $\mathcal{B}$ ($\mathcal{B}' = \mathcal{B} *_{\phi}^{\tau}$, $\pi' = \pi *_{\phi}^{\tau}$) as follows: for all interpretation $\mathcal{I}$,*

$$\pi'(\mathcal{I}) = \begin{cases} \frac{\pi(\mathcal{I})}{\Pi([\phi])}, & \text{if } \mathcal{I} \models \phi \text{ and } \mathcal{B}(\neg\phi) < 1; \\ 1, & \text{if } \mathcal{I} \models \phi \text{ and } \mathcal{B}(\neg\phi) = 1; \\ \min\{\pi(\mathcal{I}), (1-\tau)\}, & \text{if } \mathcal{I} \not\models \phi. \end{cases} \tag{6}$$

The second case in Equation 6 provides for the *revision* of beliefs that contradict ϕ. 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 6). The possibility of world situations where ϕ is true may only increase due to the first case in equation 6 or revision (second case of Equation 6). If information from a fully trusted source contradicts an existing proposition that is fully believed, then revising with the above operator leads the agent to believe the more recent information and give up the oldest to restore consistency.

It has been proved [8] that the belief change operator $*$ of Definition 8 obeys a possibilistic formulation of the AGM revision rationality postulates K$*$1–K$*$8 [12].

It is easy to verify that the $*$ operator is a generalization of the possibilistic conditioning operator of Dubois and colleagues [10].

	$\neg p$ $\neg t$	$\neg p$ t	p $\neg r$	p r $\neg t$	p r t
$\neg a, \neg b, \neg f$	1	1	0	0	0
$\neg a, \neg b, f, \neg h$	1	1	0	0	0
$\neg a, b, \neg f$	1	0.1	0	0	0
$\neg a, b, f, \neg h$	1	0.1	0	0	0
$a, \neg b, \neg f$	1	1	0	1	1
$a, \neg b, f, \neg h$	1	1	0	1	1
$a, b, \neg f$	1	0.1	0	1	0.1
$a, b, f, \neg h$	1	0.1	0	1	0.1
f, h	0	0	0	0	0

Figure 1. Dr. Gent's initial possibility distribution. Interpretations have been grouped together where possible, due to lack of space: when no literal appears for a given atom in a row or column heading, it is understood that the row or column applies for both truth assignments.

Example (continued) Dr. Gent's *a priori* beliefs may be modeled by assuming Dr. Gent at some point had no beliefs at all (π is 1 everywhere), and then was "told":

- $f \supset \neg h$ with certainty 1 (i.e., $f \supset \neg h$ by a fully trusted source),
- $b \supset \neg t$ with certainty 0.9 (i.e., $b \supset \neg t$ by a source with trust $\tau = 0.9$),
- $\neg(r \wedge a) \supset \neg p$ with certainty 1,

which yields the possibility distribution shown in Figure 1.

At this point, the following happens:

- $1/a$: Dr. Gent receives the notification of acceptance of his paper: the source is the program chair of ECAI, whom Dr. Gent trusts in full;
- $0.75/q$: soon after learning that the paper has been accepted, Dr. Flaky rushes into Dr. Gent's office to inform him that he is no more available to go to Lisbon; as always, Dr. Gent does not completely trust what Dr. Flaky tells him, as he is well-known for changing his mind very often;
- $0.2/f$: a few weeks later, Dr. Gent meets a colleague who tells him he has heard another colleague say someone on ECAI's organizing committee told her all the hotels in Lisbon are already full; Dr. Gent considers this news as yet unverified; nevertheless, he takes notice of it.

Dr. Gent's beliefs are now represented by the possibility distribution shown in Figure 2.

	$\neg p$ $\neg r$ $\neg t$ $\neg q$	$\neg p$ $\neg r$ $\neg t$ q	$\neg p$ $\neg r$ t $\neg q$	$\neg p$ $\neg r$ t q	r $\neg t$ $\neg q$	r $\neg t$ q	r t $\neg q$	r t q	p $\neg r$
$\neg a$	0	0	0	0	0	0	0	0	0
$a, \neg b, \neg f$	0.25	0.8	0.25	0.8	0.25	0.8	0.25	0.8	0
$a, \neg b, f, \neg h$	0.25	1	0.25	1	0.25	1	0.25	1	0
$a, b, \neg f$	0.25	0.8	0.1	0.1	0.25	0.8	0.1	0.1	0
$a, b, f, \neg h$	0.25	1	0.1	0.1	0.25	1	0.1	0.1	0
a, f, h	0	0	0	0	0	0	0	0	0

Figure 2. Dr. Gent's final possibility distribution. Interpretations have been grouped together where possible, due to lack of space: when no literal appears for a given atom in a row or column heading, it is understood that the row or column applies for both truth assignments.

2.6 Updating Desires

The set of the agent's justified 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*

$$\mathcal{J}(\phi) = \Delta([\phi]) = \min_{\mathcal{I} \models \phi} u(\mathcal{I}). \tag{7}$$

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 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 $R = \alpha_R \Rightarrow_D^+ \phi$, $\mathrm{Deg}(R) = \alpha_R$.*

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

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

$$u_i(\mathcal{I}) \leftarrow \begin{cases} \max_{R \in \mathcal{R}_J^{\mathcal{I}}} \mathrm{Deg}_{i-1}(R), & \text{if } \mathcal{R}_J^{\mathcal{I}} \neq \emptyset, \\ 0, & \text{otherwise,} \end{cases}$$

   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 corresponding to mental state $\mathcal{S}$.*

It has been proved [8] that Algorithm 1 always terminates and

Proposition 1 $\|\mathrm{Img}(u)\| \leq \|\mathcal{R}_J\| + 1$, *where* $\mathrm{Img}(u) = \{\alpha : \exists \mathcal{I}\ u(\mathcal{I}) = \alpha\}$.

3 Goals

We can now turn to the properties of desires and goals. In particular, it should be pointed out that, while desires may be inconsistent, goals must be defined as a consistent subset of desires. We start by giving the possibilistic version of desire specificity originally introduced by Lang in [13]. Such a notion will be used for goal election.

Proposition 2 *Let $S \subseteq \mathcal{L}$ be a set of formulas. Let $\phi_1, \phi_2 \in S$ be such that ϕ_1 is more specific than ϕ_2, i.e., $[\phi_1] \subseteq [\phi_2]$. Then,*

$$\mathcal{J}(\phi_2) \leq \mathcal{J}(\phi_1). \tag{8}$$

Proof: $[\phi_1] \subseteq [\phi_2]$ implies that $\exists X \subseteq \Omega$ such that $[\phi_1] \cup X = [\phi_2]$. From the properties of Δ (see Section 2.1), we have $\Delta([\phi_2]) = \min(\Delta([\phi_1]), \Delta(X))$. Therefore, $\mathcal{J}(\phi_2) \leq \mathcal{J}(\phi_1)$. □

Corollary 1 *If $\phi_1 \equiv \phi_2$, then,*

$$\mathcal{J}(\phi_1) = \mathcal{J}(\phi_2). \tag{9}$$

Proposition 3 *If an agent desires $\phi_1 \wedge \phi_2 \in \mathcal{L}$, this does not imply that ϕ_1 (or ϕ_2) is also necessarily desired by the agent. Formally:*

$$\mathcal{J}(\phi_1 \wedge \phi_2) > 0 \not\Rightarrow \mathcal{J}(\phi_1) > 0. \tag{10}$$

Example (continued) Dr. Gent's desires may now be determined, based on the desire-generation rules R_1, R_2, and R_3, and Dr. Gent's beliefs, represented by the possibility distribution shown in Figure 2, by applying Algorithm 1, which stops at iteration $i = 3$ with the qualitative utility distribution shown in Figure 3.

	$\neg p$ $\neg r$ $\neg t$	$\neg p$ $\neg r$ t	p $\neg r$	r
$\neg h$	0	0	0.75	1
h	0	0.75	0.75	1

Figure 3. Dr. Gent's final qualitative utility distribution. Interpretations have been grouped together where possible, due to lack of space: when no literal appears for a given atom in a row or column heading, it is understood that the row or column applies for both truth assignments.

As expected, $\mathcal{J}(r) = 1$ and $\mathcal{J}(t \wedge h) = \mathcal{J}(p) = 0.75$, but $\mathcal{J}(t) = \mathcal{J}(h) = 0$ (see Proposition 3). However, these are not all of Dr. Gent's desires. Other desires are justified under these conditions, for instance $\mathcal{J}(\neg a \wedge p) = 0.75$ and $\mathcal{J}(\neg a \wedge r) = 1$, and even $\mathcal{J}(\neg r \wedge p) = 0,75$ and $\mathcal{J}(r \wedge \neg p) = 1$. One may find some of these desires naïve or unrealizable; however, they are absolutely legitimate and rational. Indeed, who wouldn't want to present a paper without having paid for it ($\neg r \wedge p$)? Furthermore, who wouldn't want to present his/her paper, although rejected ($\neg a \wedge p$)?

In particular, for all $\phi \in \mathcal{L}$,

$$\begin{aligned} \phi \models r &\Leftrightarrow \mathcal{J}(\phi) = 1, \\ \phi \models \neg r \wedge (p \vee (t \wedge h)) &\Leftrightarrow \mathcal{J}(\phi) = 0.75, \\ \phi \models \neg p \wedge \neg r \wedge \neg(h \wedge t) &\Leftrightarrow \mathcal{J}(\phi) = 0. \end{aligned}$$

Definition 11 *The overall possibility of a set $S \subseteq \mathcal{L}$ of formulas is*

$$\Pi([S]) = \max_{\mathcal{I} \in [S]} \pi(\mathcal{I}). \tag{11}$$

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

Definition 12 *The overall qualitative utility, or justification, of a set $S \subseteq \mathcal{L}$ of formulas is*

$$\mathcal{J}(S) = \Delta([S]) = \min_{\mathcal{I} \in [S]} u(\mathcal{I}). \tag{12}$$

It follows from the properties of the minimum guaranteed possibility, that

$$\mathcal{J}(S) = \Delta([S]) = \Delta\left(\bigcap_{\phi \in S} [\phi]\right) \geq \max_{\phi \in S}\{\Delta([\phi])\} = \max_{\phi \in S}\{\mathcal{J}(\phi)\}. \tag{13}$$

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,*

$$\mathcal{J}(S \cup \{\phi\}) \geq \mathcal{J}(S); \tag{14}$$

$$\mathcal{J}(S) \geq \mathcal{J}(S \setminus \{\phi\}). \tag{15}$$

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 15 is obtainend by replacing S with $S' \setminus \{\phi\}$ in Equation 14. □

Proposition 5 *Let $S \subseteq \mathcal{L}$ be a set of formulas. Let $\phi_1, \phi_2 \in S$ be such that ϕ_1 is more specific than ϕ_2, i.e., $[\phi_1] \subseteq [\phi_2]$. Let us consider $S' = S \setminus \{\phi_1\}$ and $S'' = S \setminus \{\phi_2\}$. Then,*

$$\mathcal{J}(S') \leq \mathcal{J}(S''). \tag{16}$$

Proof: $[\phi_1] \subseteq [\phi_2]$ implies that $[\neg\phi_2] \subseteq [\neg\phi_1]$. $[S'] = [S] \cap [\neg\phi_1]$ and $[S''] = [S] \cap [\neg\phi_2]$. Therefore, $[S] \cap [\neg\phi_2] \subseteq [S] \cap [\neg\phi_1]$. Thanks to Proposition 2 we have $\mathcal{J}(S') \leq \mathcal{J}(S'')$. □

Corollary 2 *If $\phi_1 \equiv \phi_2$, we have $\mathcal{J}(S') = \mathcal{J}(S'')$. Furthermore, if $S' = S \setminus \{\phi_1\}$ or $S' = S \setminus \{\phi_2\}$, we have $\mathcal{J}(S) = \mathcal{J}(S')$.*

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, 7, 16].

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 takes 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, $\Pi([S]) = 0$.

Proposition 6 *Let $S_1, S_2 \subseteq \mathcal{L}$ be sets of desire formulas such that S_1 is more specific than S_2, i.e., $[S_1] \subseteq [S_2]$. Then,*

$$\Pi([S_1]) \leq \Pi([S_2]), \quad (17)$$
$$\mathcal{J}(S_2) \leq \mathcal{J}(S_1). \quad (18)$$

Proof: $[S_1] \subseteq [S_2]$ implies that $\exists X \subseteq \Omega$ such that $[S_1] \cup X = [S_2]$. From the properties of Π and Δ described in Section 2, we have $\Pi([S_2]) = \max(\Pi([S_1]), \Pi(X))$ and $\mathcal{J}(S_2) = \Delta([S_2]) = \min(\Delta([S_1]), \Delta(X))$ therefore $\Pi([S_1]) \leq \Pi([S_2])$ and $\mathcal{J}(S_2) \geq \mathcal{J}(S_1)$. □

Corollary 3 *If S_1 is more specific than S_2 and S_2 is logically inconsistent, then S_1 is logically inconsistent.*

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

Definition 13 *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 14 (Goal set) *The γ-possible goal set is*

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

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

$$G^* = G_{\gamma^*}, \quad \gamma^* = \max_{G_\gamma \neq \emptyset} \gamma. \quad (19)$$

Our agents are thus supposed to be pragmatic and risk-averse. Their policy is not just to choose the most justified set of desires as goals, but the most justified among the most possible ones. This way, the goal set may not correspond to the set with higher utility; however, it will correspond to the set with the highest possibility to be fulfilled.

Let $\mathrm{Img}(\pi)$ be the level set of possibility distribution π and $\mathrm{Img}(u)$ be the level set of qualitative distribution u. Notice that $\mathrm{Img}(u)$ is finite by Proposition 1; $\mathrm{Img}(\pi)$ is also finite, independently of Ω being finite, if we assume every agent to be created with a zero-knowledge possibility distribution π_0 such that $\pi_0(\mathcal{I}) = 1$ for all $\mathcal{I} \in \Omega$ and to have a finite history of belief changes, a very reasonable assumption indeed.

By proposition 6, less specific desires have higher possibility to be fulfilled; besides, more specific desires cannot be preferred less. Therefore, if one manages to find the least specific desire with the highest utility for a given level of possibility, it would be a loss of time to look for more specific desires. This is the basic idea behind the following two algorithms, that allow an agent to compute G_γ for a given possibility lower bound γ, and the optimal goal set G^*.

Algorithm 2 (Computing G_γ for a given γ)

1. $\delta \leftarrow \max \mathrm{Img}(u)$;
2. *determine the least specific formula ϕ such that $\mathcal{J}(\phi) \geq \delta$ as follows:*
$$\phi \leftarrow \bigvee_{u(\mathcal{I}) \geq \delta} \phi_{\mathcal{I}},$$
where $\phi_{\mathcal{I}}$ denotes the minterm *of $\mathcal{I}$, i.e., the formula satisfied by $\mathcal{I}$ only;*
3. *if $\Pi([\phi]) \geq \gamma$, terminate with $G_\gamma = \{\phi\}$; othervise,*
4. $\delta \leftarrow \max\{\alpha \in \mathrm{Img}(u) : \alpha < \delta\}$, *0 if no such α exists;*
5. *if $\delta > 0$, go back to Step 2;*
6. *terminate with $G_\gamma = \emptyset$.*

Algorithm 3 (Goal Election)

1. $\gamma \leftarrow \max \mathrm{Img}(\pi) = 1$, *since π is normalized;*
2. *compute G_γ by Algorithm 2;*
3. *if $G_\gamma \neq \emptyset$, terminate with $\gamma^* = \gamma$, $G^* = G_\gamma$; otherwise,*
4. $\gamma \leftarrow \max\{\alpha \in \mathrm{Img}(\pi) : \alpha < \gamma\}$, *0 if no such α exists;*
5. *if $\gamma > 0$, go back to Step 2;*
6. *terminate with $G^* = \emptyset$: no goal may be elected.*

Example (continued) To determine a consistent set of goals to commit to, Dr. Gent must perform a goal election. Applying Algorithm 3 in this case yields $\gamma^* = 1$ and $G^* = \{r\}$: Dr. Gent must pay the registration, that is for sure; planning his trip is less urgent, for Dr. Flaky, as far as Dr. Gent believes, might still change his mind.

4 Belief-Goal Interaction

Bratman pointed out in [4] that "we can consider irrational for an agent to intend to do an act A and at the same time believe that it will not do A. However, it is rational for the agent to intend to do A and not believe that it will do A". More precisely, it is irrational for an agent to have beliefs which are inconsistent with its intentions, while it is rational to have incomplete beliefs about its intentions. For Bratman, these two principles, *intention-belief consistency* and *intention-belief incompleteness*, are referred to as the *asymmetry thesis*.

Rao and Georgeff [17] claim that the asymmetry thesis can be extended to the relationship between intentions and goals, and goals and beliefs. Here, we will show that the proposed possibilistic formalism for representing a rational agent obeys the two principles of belief-goal consistency and belief-goal incompleteness.

4.1 Asymmetry Thesis

The two principles of the asymmetry thesis corresponding to the belief-goal relation — the avoidance of belief-goal inconsistency ([BG-ICN]) and the belief-goal incompleteness ([BG-ICM]) — are stated as follows in our possibilistic-based formalism:

Theorem 1 (Avoidance of belief-goal inconsistency (BG-ICN)) *If the agent adopts ϕ as a goal (at a given level of possibility γ^*), then the agent should not (somehow) believe $\neg\phi$:*

$$\phi \in G^* \Rightarrow \mathcal{B}(\neg\phi) \leq 1 - \gamma^*. \tag{20}$$

Proof: $\phi \in G^* \Rightarrow \phi \in S^* : \Pi([S^*]) \geq \gamma^*$; by the properties of Π,

$$\begin{aligned} \Pi([S^*]) = \Pi(\bigcap_{\phi \in S^*} [\phi]) &\leq \min_{\phi \in S^*} \Pi([\phi]) \\ \gamma^* &\leq \Pi([\phi]) \Rightarrow \mathcal{B}(\neg\phi) \leq 1 - \gamma^*. \end{aligned}$$

□

Theorem 2 (Belief-goal incompleteness (BG-ICM)) *The agent can adopt ϕ as a goal (at a given level of possibility γ^*), even though it does not (somehow) believe ϕ:*

$$\phi \in G^* \not\Rightarrow \mathcal{B}(\phi) > 1 - \gamma^*. \tag{21}$$

Proof: We must prove that $\exists\gamma^* : (\phi \in G^*) \wedge (\mathcal{B}(\phi) \leq 1 - \gamma^*)$. If γ^* is such that $0 \neq \gamma^* \leq \Pi([\phi]) = \alpha < 1$, $\Pi([\neg\phi]) = 1$ (because $\max(\Pi([\phi]), \Pi([\neg\phi]))$=1). Therefore, $\mathcal{B}(\phi) = 0 \leq 1 - \gamma^*$. □

The way in which the relationships between beliefs and goals are captured can have a significant impact on the design of a rational agent [15]. In particular, it can lead to the *side-effect problem* and to the *transference problem*. In the following, we show that our formalism avoids such problems.

4.2 Side-Effect-Free Principle

The possibilistic counterpart of the Belief-Goal side-effect-free principle states that,

Theorem 3 *If an agent adopts ϕ as a goal, no matter how strongly it believes $\phi \supset \psi$, it should not be forced to adopt ψ as a goal as a side-effect. Formally:*

$$\phi \in G^* \wedge \mathcal{B}(\phi \supset \psi) > 1 - \gamma^* \not\Rightarrow \psi \in G^*. \tag{22}$$

Proof: We have to prove that there exists a mental state $\mathcal{S}$ such that $\phi \in G^*$ and $\mathcal{B}(\phi \supset \psi) > 1 - \gamma^*$, but $\psi \notin G^*$. Let $\mathcal{A} = \{\phi, \psi\}$: we construct such an $\mathcal{S}$ by letting

π	$\neg\phi$	ϕ
$\neg\psi$	x	0
ψ	y	1

, $\mathcal{R}_J = \{1 \Rightarrow^+_D \phi\}$, $\rightsquigarrow$

u	$\neg\phi$	ϕ
$\neg\psi$	0	1
ψ	0	1

,

where x and y may be any possibility degrees in $[0, 1]$. Now, applying Algorithm 3 yields $\gamma^* = 1$ and $G^* = \{\phi\}$. Therefore, $\phi \in G^*$ and $\mathcal{B}(\phi \supset \psi) = 1 > 1 - \gamma^* = 0$, but $\psi \notin G^*$. □

An intuitive understanding of the above counterexample is that an agent has some problems with its tooth. Even though it would like to get its tooth fixed (ϕ), it would not like to feel the pain (ψ) that is believed to inevitably follow.

4.3 Non-Transference Principle

The possibilistic counterpart of the non-transference principle can be stated as:

Theorem 4 *No matter how strongly an agent believes in a proposition ϕ, it should not be forced to adopt ϕ as a goal. Formally:*

$$\mathcal{B}(\phi) > 1 - \gamma^* \not\Rightarrow \phi \in G^*. \tag{23}$$

Proof: It is enough to consider a formula ϕ such that $\mathcal{B}(\phi) > 1 - \gamma^*$ and $\mathcal{J}(\phi) = 0$, that is, $\nexists\mathcal{I} \models \phi\ u(\mathcal{I}) > 0$. □

5 Conclusion

The properties of a full-fledged theoretical framework for goal generation in BDI agents have been investigated.

The formalism obeys two important principles of agent rationality, namely the principles of belief-goal consistency and belief-goal incompleteness. Furthermore, it has been proved that, while obeying these principles, the formalism avoids two well-known design pitfalls, namely the side-effect problem and the transference problem.

Therefore, a formal framework based on possibility theory, like the one we have investigated in this paper, should be taken into account as a serious candidate for a theoretical foundation of cognitive agents that deal with uncertain information and partial trust.

Here we have defined belief-goal consistency in an atemporal propositional setting through the hypothesis that beliefs and goals can refer to situations in the past, present, and future. We plan to generalize our formalism in order to explicitly deal with time.

REFERENCES

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

[2] 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).

[3] J. Blee, D. Billington, and A. Sattar, 'Reasoning with levels of modalities in BDI logic', in *Proceedings of PRIMA'07*, pp. 410–415. Springer-Verlag, (2009).

[4] M. Bratman, *Intentions, Plans, and Practical Reason*, Harvard University Press, Cambridge, 1987.

[5] A. Casali, L. Godo, and C. Sierra, 'Graded BDI models for agent architectures', in *CLIMA V*, pp. 18–33, (2004).

[6] C. Castelfranchi and F. Paglieri, 'On the integration of goal dynamics and belief structures', *Synthese*, (2007).

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

[8] C. da Costa Pereira and A. Tettamanzi, 'An integrated possibilistic framework for goal generation in cognitive agents', in *Proceedings of AAMAS'10*, pp. 1239–1246. IFAAMAS, (2010).

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

[10] 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).

[11] 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).

[12] P. Gärdenfors, 'Belief revision: A vademecum', in *Meta-Programming in Logic*, 1–10, Springer, Berlin, (1992).

[13] J. Lang, 'Conditional desires and utilities: an alternative logical approach to qualitative decision theory', in *Proceedings of ECAI'96*, pp. 318–322, (1996).

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

[15] A. S. Rao and M. P. Georgeff, 'Decision procedures for BDI logics', *J Logic Computation*, **8**(3), 293–343, (June 1998).

[16] A.S. Rao and M.P. Georgeff, 'Asymmetry thesis and side-effect problems in linear-time and branching-time intention logics', in *Proceedings of IJCAI'91*, pp. 498–505, (1991).

[17] A.S. Rao and M.P. Georgeff, 'Modeling rational agents within a BDI-architecture', in *Proceedings of KR'91*, pp. 473–484, (1991).

[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).